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.
An experimental analysis of skill responding in high school physical education

Ward, Phillip Charles, Ph.D.
The Ohio State University, 1993
AN EXPERIMENTAL ANALYSIS OF SKILL RESPONDING
IN HIGH SCHOOL PHYSICAL EDUCATION

DISSERTATION

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

by


* * * * *

The Ohio State University
1993

Dissertation Committee:
Dr. Daryl Siedentop
Dr. Mary O'Sullivan
Dr. John Cooper

Approved by
Advisor
School of Health, Physical Education and Recreation
To my father

Phillip Robert Ward
ACKNOWLEDGMENTS

Much of a dissertation is about recognizing conceptual and methodological debts. A different kind of debt is owed to those who helped shape my behavior through graduate school.

If graduate school is considered a treatment package, then my advisor Daryl Siedentop as a component, has had a robust effect upon my behavior. He let me pursue my academic interests, and at times suggested particular directions to follow. He has helped me to grapple with the big picture, and defined for me, by example, scholarly conduct - -Thank you.

My gratitude is especially due to Mary O’Sullivan who challenged me to look beyond my own beliefs and values, who changed the way I viewed teaching supervision, who has by example been a model of professional commitment, and who never turned me away when I knocked on her office door, with a question or two.

John Cooper, Bill Heward and Tim Heron have been my “second” faculty - Thank you for treating me as one of your own.

To Debra Tannehill, and Sandy Stroot my sincere thanks for your advice, and guidance throughout my time at OSU.
All of us need professional colleagues, Gregg Drevno, Panayiotis Doutis, and Niki Tsangaridou have fulfilled that role for me.

Special thanks are due to Mrs. East and Mrs. West who let me enter their classrooms, who collaborated with me in the true sense of the word, and who shared in my enthusiasm for this study. Thanks are also due to Emyr Williams, Barbara Nagy, and Daniel Bullara for helping in the data collection and IOA’s.

Finally my deepest gratitude is due to Marie, in her special roles as colleague, and friend.
VITA

September 14, 1957 ...... Born - Brisbane, Australia

1979 - 1981 ............... Diploma in Teaching, North Brisbane College of Advanced Education, Australia

1981 - 1984 ............... Coaching Director - Queensland Academy of Artistic Gymnastics, Brisbane, Australia

1984 - 1987 ............... Physical Education Teacher - Manningham Road Primary School, Melbourne, Australia

1986 - 1988 ............... Graduate Diploma in Sports Science, Department of Physical Education, Victoria College, Melbourne, Australia

1987 - 1989 ............... Bachelor of Education, Department of Teacher Education, Deakin University, Geelong, Australia

1987 - 1991 ............... Consultant for the Victorian Education Department - Physical Education, Melbourne, Australia

1990 - 1991 ............... Master of Education, Department of Curriculum and Teaching, Victoria College, Melbourne, Australia

1990 - 1993 ............... Graduate Teaching Assistant - School of Health, Physical Education, & Recreation, The Ohio State University, Columbus, Ohio
PUBLICATIONS


FIELDS OF STUDY

Major Field: Health, Physical Education, and Recreation

Dr. Daryl Siedentop, Dr. Mary O'Sullivan, and Dr. Debra Tannehill

Minor Field: Applied Behavior Analysis

Dr. Daryl Siedentop, Dr. John Cooper, and Dr. William Heward
# TABLE OF CONTENTS

DEDICATION ....................................................................................................... ii

ACKNOWLEDGMENTS .................................................................................... iii

VITA .................................................................................................................. v

LIST OF TABLES ........................................................................................... ix

LIST OF FIGURES ........................................................................................... xi

CHAPTER

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

   Statement of the Problem ........................................ 6
   A Framework for the Study ..................................... 9
   The Experimental Strategy ....................................... 13
   Experimental Questions ............................................ 14
   Significance of the Study .......................................... 15
   Anthropological Assumptions ................................ 16
   Limitations of the Study ............................................ 20
   Definition of Terms ...................................................... 21

II. REVIEW OF LITERATURE ............................................... 25

   Rate of Response ........................................................... 26
   Response as a Unit of Analysis in Education .. 28
   Empirical Evidence of Responding in Education .......... 32
   Instructional Arrangements Influencing Student Responding .......... 41
   A Summary of Responding in Education .......... 44
   The Responding Literature in Physical Education .......... 46
   Rate of Response as a Unit of Analysis in Physical Education .......... 48

vii
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean percent responding for high and low skilled target students across classes (Son, 1989)</td>
<td>57</td>
</tr>
<tr>
<td>2. Mean response rate per minute for more and less skilled target students across classes (Lund, 1990)</td>
<td>61</td>
</tr>
<tr>
<td>3. Mean response rate per lesson for more and less skilled target students across classes (Lund, 1990)</td>
<td>61</td>
</tr>
<tr>
<td>4. Mean number of opportunities to respond and criterion trials by classes in Dugas' (1983) study</td>
<td>67</td>
</tr>
<tr>
<td>5. Design arrangement across sessions for the first setting</td>
<td>139</td>
</tr>
<tr>
<td>6. Interobserver agreement measures for the dependent variables</td>
<td>150</td>
</tr>
<tr>
<td>7. Background information of target students in Class 1</td>
<td>156</td>
</tr>
<tr>
<td>8. Ms. East's task statements during the basketball study</td>
<td>157</td>
</tr>
<tr>
<td>9. Means and ranges of correct responses for each task across treatments for each target student in Class 1. (Basketball)</td>
<td>162</td>
</tr>
<tr>
<td>10. Background information of target students in Class 2</td>
<td>170</td>
</tr>
</tbody>
</table>
11. Mrs. West's task statements during the basketball study ............................................................... 172

12. Means and ranges of correct responses for each task across treatments for each target student in Class 2 (Basketball) .................................................. 175

13. Means and ranges of correct responses for each task across treatments for each target student in Class 1. (Volleyball) .................................................. 186

14. Means of incorrect responses for each task across baseline and treatment conditions for each target student in Class 1. (Volleyball) ......................... 194

15. Means and ranges of correct responses for each task across treatments for each target student in Class 2. (Volleyball) ................................................. 202

16. Means of incorrect responses for each task across baseline and treatment conditions for each target student in Class 2. (Volleyball) ......................... 203

17. Means and ranges of correct responses for each task across treatments for each target student in Class 1. (Badminton) ........................................... 222

18. Means of incorrect responses for each task across baseline and treatment conditions for each target student in Class 1. (Badminton) ......................... 222

19. Means and ranges of correct responses for each task across treatments for each target student in Class 2. (Badminton) ........................................... 234

20. Means of incorrect responses for each task across baseline and treatment conditions for each target student in Class 2. (Badminton) ......................... 234

21. The relationship between tasks and goals ........ 270
LIST OF FIGURES

FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Morningside Model of Generative Instruction</td>
<td>40</td>
</tr>
<tr>
<td>2. Rate per minute of total responses and successful responses for high skilled and low skilled students in Graham’s (1986) study.</td>
<td>52</td>
</tr>
<tr>
<td>3. Percentage of adversities responded to favorably by the target team as a function of the rate of reinforcement (Mace et al., 1992).</td>
<td>85</td>
</tr>
<tr>
<td>4. A model for the analysis of instructional task systems (Son, 1989)</td>
<td>108</td>
</tr>
<tr>
<td>5. The categories used in the analysis of the instructional contingencies</td>
<td>110</td>
</tr>
<tr>
<td>6. An example of a completed coding sheet</td>
<td>121</td>
</tr>
<tr>
<td>7. An example of a peer recording sheet</td>
<td>130</td>
</tr>
<tr>
<td>8. A model of the proposed alternating treatment design.</td>
<td>138</td>
</tr>
<tr>
<td>9. Percentage of class time spent on lesson content for Class 1 during the basketball study</td>
<td>158</td>
</tr>
<tr>
<td>10. Correct responses per minute for each target student in Class 1. (Basketball)</td>
<td>159</td>
</tr>
<tr>
<td>11. Percentage of class time spent on lesson content for Class 2 during the basketball study</td>
<td>173</td>
</tr>
<tr>
<td>12. Correct responses per minute for each target student in Class 2. (Basketball)</td>
<td>174</td>
</tr>
</tbody>
</table>
13. Percentage of class time spent on lesson content for Class 2 during the volleyball study .......... 188

14. Correct responses per minute for target students in Class 1 for Sprints and Peer Sprints across instructional tasks (Volleyball) .................... 189

15. Correct and incorrect responses for the wall drill for target students in Class 1 during the volleyball study ................................................. 190

16. Correct and incorrect responses for the fed drill for target students in Class 1 during the volleyball study ................................................. 191

17. Correct and incorrect responses for the 1 V 1 drill for target students in Class 1 during the volleyball study ................................................. 192

18. Percentage of team responses made by target students in Class 1 and number of team responses for students during the volleyball study .......... 193

19. Percentage of class time spent on lesson content for Class 2 during the volleyball study ........ 204

20. Correct responses per minute for target students in Class 2 for Sprints and Peer Sprints across instructional tasks (Volleyball) .................... 205

21. Correct and incorrect responses for the wall drill for target students in Class 2 during the volleyball study ................................................. 206

22. Correct and incorrect responses for the fed drill for target students in Class 2 during the volleyball study ................................................. 207

23. Correct and incorrect responses for the 1 V 1 drill for target students in Class 2 during the volleyball study ................................................. 208
24. Percentage of team responses made by target students in Class 2 and number of team responses for students during the volleyball study ............. 209

25. Percentage of class time spent on lesson content for Class 1 during the badminton study ............. 223

26. Correct responses per minute for target students in Class 1 for Sprints and Peer Sprints across instructional tasks (Badminton) ....................... 224

27. Correct and incorrect responses for the fed drill for target students in Class 2 during the badminton study ................................................................. 225

28. Correct and incorrect responses for the 1 V 1 drill for target students in Class 2 during the badminton study ................................................................. 226

29. Percentage of team responses made by target students in Class 1 and number of team responses for students during the badminton study ............. 227

30. Percentage of class time spent on lesson content for Class 2 during the badminton study ............. 235

31. Correct responses per minute for target students in Class 2 for Sprints and Peer Sprints across instructional tasks (Badminton) ....................... 236

32. Correct and incorrect responses for the fed drill for target students in Class 2 during the badminton study ................................................................. 237

33. Correct and incorrect responses for the 1 V 1 drill for target students in Class 2 during the badminton study ................................................................. 238

34. Percentage of team responses made by target students in Class 2 and number of team responses for students during the badminton study. ............. 239
The public school was invented to bring the services of a private tutor to more than one student at a time. As the number of students increased, each one necessarily received less attention. By the time the number had reached 25 or 30, personal attention could be sporadic at best. Textbooks were invented to take over some of the work of the tutor, but they could not do two important things. They could not, as the tutor did, immediately evaluate what each student said, nor could they tell the student exactly what to do next. (Skinner, 1989, pp. 85-86)

Skinner continued by noting that despite the change in the role of the public school “American education turns out some successful products. But not enough of them” (Skinner, 1989, pp. 85-86). The products of education often serve as the focus for calls for educational reform. If the educational establishment was a beach then educational reform might be described as a series of waves. Each wave shaping the beach in one direction or another. Each wave of educational reform is the product of
heightened public awareness of schools and a discontent with what is found. This discontent produces calls for change. Over the past decade there have been two significant "waves" of educational reform. The first wave began with a series of reform reports (Adler, 1982; Boyer, 1983; Education Commission of the States, 1983; National Science foundation, 1983; Sizer, 1984) and was epitomized by the release of the publication "A nation at risk: The imperative for educational reform" (National Commission on Excellence in Education, 1983). The National Commission on Excellence in Education charged that America is at risk "because competitors throughout the world are overtaking our once unchallenged lead in commerce, industry, science and technological innovation" (Goldberg & Harvey, 1983 p. 15). Much of the responsibility for America's "at risk" status has been laid firmly in the lap of education (National Commission on Excellence in Education, 1983).

A second wave of reform is currently transpiring and is effervescing with solutions to American educational problems including national curricula (Finn, 1991) school restructuring (Murphy, 1992); arguments for a marketplace educational system (Chubb & Moe, 1990); and calls for changes in school and teacher accountability (Peterson, 1992).

Whatever the focus for educational reform, inevitably school-based change is reduced to the dynamic relationship between the instruction and the student in a specific context. This relationship though capable of being influenced by several
other relations (e.g., teachers, principals, district policies, and a community's socio-economic status) doesn't change the basic fact that if students improve it will be because of instruction provided or arranged in an educational setting (e.g., teachers, computers, tutors). This position (of improving the instructional conditions) is an explicit assumption of much of the current reform literature which views the "student as worker" and supports efforts which improve the conditions for student work.

Among other things, implementing reform requires finance. The costs and benefits of reforms are not only to be judged in terms of the impact upon a student, but also the impact upon the economy (be it local, state or federal). An effective reform is not only one which is necessary, but one which is cost efficient. Improving the effectiveness of teachers makes good educational and economic sense as an achievable reform. As Medley (1979) observed:

The effect of schooling on the individual pupil depends to a considerable extent on who his [or her] teacher is . . . Personnel costs themselves represent so large a share of the day to day cost of education that the best hope for improvement in cost-effectiveness lies in improving the effectiveness of teachers (p. 11).

The search for improving the effectiveness of teachers began in the early 1970's. Persuasive reviews by Rosenshine and Furst (1973), and Dunkin and Biddle (1974) were the precursors to a decade and a half of extensive, large scale
correlational and experimental research studies focused on defining and identifying the variables of teaching effectiveness. During this time one strand of teaching effectiveness research, the process-product paradigm (Dunkin and Biddle, 1974), has evolved from a macro-analytic focus on teaching behaviors and end of year student achievement, to a mediating process model where the search was for a variable or variables that would allow an immediate assessment of the effects of teaching rather than waiting for the more distant, end of year achievement test (e.g., Academic learning time, Berliner, 1979). This led to a micro-analytic focus on the reciprocal and interactive nature of teacher and student behaviors occurring within a lesson. This interactive view of teaching has become known as the ecological paradigm (Doyle, 1977). Two voluminous “Handbooks of research on teaching” (Travers, 1973; Wittrock, 1986) each produced at the end of a decade of study testify to the efficacy and the extent of teaching effectiveness research. The process-product paradigm and its progeny (the mediating process paradigm and the ecological paradigm) have furnished utilitarian (and economical) findings using student achievement as a dependent variable.

The process-product paradigm was not the only paradigm in which research on teaching effectiveness was conducted and not all paradigms have defined effectiveness in terms of student achievement. For example, Shulman (1986) reviewed five research paradigms each of which views teaching
effectiveness differently from the other. Findings from these other paradigms have provided alternatives to viewing teaching effectiveness other than in terms of student achievement. Notwithstanding these alternatives, student achievement remains an important benchmark against which to judge the efficacy of teachers and treatments.

Research in regular education and special education classrooms over the past 30 years has, using student achievement as a dependent variable, yielded a number of effective practices or technologies such as mastery learning (Anderson & Burns, 1987; Bloom, 1984; Guskey & Gates, 1988; Kulik, Kulik, & Bangert-Downs, 1990); tutoring (Bloom, 1984; Cohen, 1986); classwide peer tutoring (Greenwood, Delquadri, & Carta, 1988, Greenwood, Delquadri, & Hall, 1989; Greenwood, Terry, Arreaga-Mayer, & Finney, 1992); instructional alignment (Cohen, 1987); personalized systems of instruction, (Keller & Sherman, 1974; Buskist, Cush, & DeGrandpre 1991; Sherman, 1992); direct instruction (Rosenshine, 1979); Direct Instruction (Becker & Carnine, 1980, Kinder & Carnine, 1991); and programmed instruction (Skinner, 1968, 1989, Vargas, & Vargas, 1991). Effective practices have also included a curricular decision making technique called precision teaching (Lindsley, 1990, 1992; White, 1986).

In physical education, research programs (Metzler, 1982; 1986; 1989; Siedentop, 1982; 1983) were adopted to answer teaching effectiveness research questions in a similar manner
to those asked by classroom researchers. Pieron (1992) has recently summarized 22 experimental studies in physical education divided into four categories: research methods, student and teacher behaviors related to learning, the student as the unit of analysis, and the specificity of teaching effectiveness. Findings from these studies closely parallel the findings from the classroom process-product literature (Harrison, 1987; Locke, 1982; Metzler, 1989; Pieron, 1992; Siedentop, 1982, 1991). The assumption these studies were based on was stated by Locke (1979) who noted that despite the ecological differences between gymnasia and classroom settings, effective instructional techniques are similar and that which works in the classroom, will also work in the gymnasium. This assumption has produced similar research questions, similar research methods and physical education-specific observation instruments.

Statement of the Problem

In 1984 Bloom reported on two completed dissertations that compared learning under three conditions: conventional instruction, mastery learning and tutoring. Conventional instruction was characterized by a 1:30 teacher student ratio where tests were given periodically to measure student achievement. Mastery learning involved providing the same tests as those used in the conventional instruction group in the same class conditions (i.e., 1:30 teacher student ratio). In this
condition errors made by students who took the tests served as the focus for corrective techniques and further parallel tests. The goal of this condition was mastery of the content presented. Mastery learning used a deficiency based model where the focus was on remediation. The third condition, tutoring, was significantly different in its instructional conditions. In the tutorial conditions the tutor worked one-on-one or at most one-on-three with students using the following procedure: instruction-test-feedback-corrective procedures and parallel (similar format and content) formative tests. In reporting the results Bloom (1984) noted “the most striking of the findings is that under the best learning conditions we can devise (tutoring), the average student is 2 sigma above the average control student taught under conventional group methods of instruction” (p. 4). (The standard deviation or sigma is a measure of group dispersion from the mean). Under conventional instruction around 68% of the population occupy one standard deviation either side of the mean. If a form of instruction shifted the mean higher and kept the variance low this would indicate that more students had higher achievement than would be normally predicted using conventional instruction. This is exactly what happens in tutoring. The effect of tutorial instruction is to shift the mean higher by two standard deviations (i.e., 2 sigma) and reduce variance. The average student in a tutorial instruction class performed better than 98% of the students in the conventional instruction class (Bloom, 1984).
These results have been replicated sufficiently for teachers and researchers to have confidence in the validity and generalizability of the findings (Bloom, 1984; Kulik et al., 1990, Greenwood, et al., 1988, Greenwood, et al., 1989; Greenwood, et al., 1992). As a result of this research Bloom (1984) posed what he called the "2 sigma problem." "Can researchers and teachers devise teaching-learning conditions that will enable the majority of students under group instruction to attain levels of achievement that can at present be reached only under good tutoring conditions"? (Bloom, p. 4)

Classroom research in both regular and special education classrooms has generated a number of effective teaching strategies that produce mastery using high response rates of correct performance (e.g., Direct Instruction, precision teaching, tutoring, mastery learning). While teaching effectiveness research in physical education has adopted the assumptions, questions, and methods of its classroom cousins it has not yet pursued the question "What are the physical education analogs to the effective instructional practices used in classrooms that produce mastery using high response rates of correct performance"? Answering this question cannot be achieved using descriptive-analytic or correlational research paradigms. It can be answered only by an experimental paradigm where the controlling relations can be elucidated and where instructional strategies can be compared to provide an answer.
A Framework for the Study

Bloom (1984) argued that the solution to the “2 sigma problem” lay in the search for what he called “alterable variables.” Among the most productive of these searches conducted by classroom researchers was reported by Cohen (1987) “We decided to seek a magic bullet—the most potent variable among many underlying mastery learning” (p. 17). Cohen’s alterable variable was called “instructional alignment.” It described the congruence between the goals of instruction, the instruction, and the instructional assessment (Cohen, 1987). Like Bloom (1984) in reporting the effects of mastery learning, Cohen (1987) reported a 1-3 sigma shift in student achievement when goals, instructional stimuli and assessment match. The presence of instructional alignment produced an effect “about four times what we see in regular classrooms” (Cohen, 1987, p. 19).

Among the most oft cited alterable variables in research on teaching is time (Berliner, 1979; Bloom, 1984; Carroll, 1963; Rosenshine & Stevens, 1986; Walberg, 1984, ). However, it is not time itself that is being referred to, but events or responses that happen in time. Responding has been a variable that is often implied in much of the literature on time in the teaching effectiveness research. The amount or rate of responding has been investigated both directly and indirectly in teaching effectiveness research. For example, Kounin (1970) used “work involvement” as his dependent variable and found that two variables which he named “smoothness” (the absence of breaks
and stops in a lesson) and "momentum" (the pace of the lesson) correlated strongly with it. Of the two variables, the "momentum" of the lesson was the more significant variable. Kounin (1970) commented "It is more important to maintain momentum by avoiding actions that slowdown forward movement than it is to maintain smoothness by avoiding sudden starts and stops" (p. 108). Though "momentum" referred specifically to teacher behavior and to lesson management, the implication is that the students would be "pulled along" by the pace of the lesson (Doyle, 1986). Merrit (1982) has called this "pulling along" effect a "vector of activity." Rosenshine and Stevens (1986) have been more explicit in identifying rate of responding as a characteristic of effective instruction "It is exciting to watch the class or group move at a rapid pace and to watch all the students giving the correct response rapidly and confidently" (p. 380).

High rates of correct responding are a common feature of sport settings. In their study of basketball coach John Wooden, Tharp and Galimore (1976) reported on the importance of hustles and prompts that were used to intensify practice. They commented "If such concentrated effort ever finds its way into the ordinary classroom, the results might be rewarding" (p. 78). Mace et al., (1988) described a concept known as "behavioral momentum" which is the tendency for behavior to persist following a change in environmental conditions. Mace et al., (1988) found that clients in a direct care facility were more
likely to comply with requests for low-probability behavior if these requests were preceded by a period of a high rate of responding for high probability behaviors. Recently Mace, Lalli, Shea, and Nevin (1992) illustrated the effects of “behavioral momentum” in sport settings by analyzing data collected from National Collegiate Athletic Association tournament games. Mace et al., (1992) reported that the team whose members collectively had high rates of successful responding were able to pick up the pace and persevere after adversity (e.g., missed shots, fouls, unfavorable turnovers). They noted that the standard strategy used by the coaches of the opposing team to decrease the behavioral momentum was to reduce the rate of reinforcement that the other team could get by calling a “time out.”

Behavior analysts have used rate of response as their fundamental datum since 1938 when Skinner first introduced it to the natural science of behavior (Skinner, 1938). In 1950 he suggested that “Rate of responding appears to be the only datum which varies significantly and in the expected direction under conditions which are relevant to the learning process” (Skinner, 1982, p. 48). Rate of response is a good unit of measurement with which to investigate the effects of teaching. It allows data collected over different durations to be compared to a common scale (e.g., rate per minute); and it allows a measure of relative change (celeration) to be obtained. Lindsley (1991a) has asserted that rate is a dimension of behavior. His proverb “Change the frequency [rate], change the behavior” (Lindsley, 1991a, p. 5)
places particular importance on the rate of response that has not been reported elsewhere. His emphasis has been on increasing rates of correct responding as a mechanism to improve mastery of educational tasks.

In summary, there appears substantial evidence to qualify rate of responding as an alterable variable and to merit investigation into its controlling relations. This study sought to further extend the investigation into Blooms' "2 sigma problem" by investigating treatments that could increase rates of responding in high school physical education relative to normal gymnasium conditions. High school physical education is not a place that researchers and teachers associate with high levels of responding (successful or otherwise). After a year long study of physical education in 11 high schools Siedentop and O'Sullivan (1992) commented, "We found instructional settings that were at best, distinctly causal, with substantive learning gains modest. In the classes we found clear winners and losers, most frequently definable by gender and skill level" (p. 285). Many teachers are not convinced that rates of responding can be improved. A teacher interviewed in that study previously cited by Siedentop and O'Sullivan (1992) commented "I just don't believe you can push the average Joe." If rate of responding proves to be an alterable variable in physical education classes, then treatments/tasks that increase rate of responding might be a key to achieving instructional practices that produce mastery.
The Experimental Strategy

The goal of this study was to experimentally analyze rate of responding. Since rate of response provides little information about the qualitative dimensions of the response, judgments about the correct or incorrect topography of a response, and whether the performance outcome was successful or unsuccessful have also been made. Thus the dependent variables of this study were rate of response, correct and incorrect response topography and success or unsuccessful response outcome.

The experimental manipulation of rate of responding was achieved through the use of three treatments derived from treatments identified in the research on teaching literature, and which have been used in regular and special education classrooms. These treatments share a “common thread;” in that they all produce high rates of correct responding. The treatments were designed to produce high rates of correct responding when compared to standard or normal practice conditions occurring in the classes observed.

Since this is the first investigation on increasing rates of responding in high school physical education, and the first experimental analysis of tasks in physical education, the investigator sought the freedom to be “led by the data” and to modify and change the experimental conditions based on data obtained as the investigation progressed.
Experimental Questions

1. What are the effects of modified classroom instructional treatments in physical education?

1.1 Do the treatments improve the response rate, success rate, and correct topography rate of the target students relative to "standard" practice?

1.2 Are any of the treatments superior to the others relative to response rate, success rate, and correct topography rate?

1.3 In a forced choice between treatments which treatment is selected by the students?

1.4 Which (if any) of the treatments are used by the teachers in their other classes?

2. How does an increased rate of responding affect the behavior of students and teachers in physical education lessons?

2.11 Does an increased rate of responding affect high and low skilled students differently?

2.12 Does an increased rate of responding affect male and female students differently?

2.2 Does the explicitness of tasks stated by the teacher change as the rates of responding changes?

2.3 In what ways does an increased rate of responding affect how students and teachers spend their time in physical education lessons?
Significance of the Study

A task of science is to find orderly relations among phenomena (Sidman, 1960). Science does not and should not consist solely of the cataloging of facts, but rather seeks understanding by controlling and manipulating variables, the purpose of which is the production of findings of utility to society.

An important task for the researcher is to capture “good” (accurate, reliable, general, and scientifically important) data (Sidman, 1960). Science is extended by the similarities among data, not by their differences. This is achieved through programmatic research efforts where the independent variables are replicated and the effects reproduced, and by a view which, while not entirely inductive nonetheless requires researchers to be led by data that reveal themselves under experimental control.

Accordingly the significance of this study lies first with its ability to report “good” data and secondly with its contribution to science in terms of its programmatic significance. The use of the strategies and tactics of the natural science of behavior to record and represent the events observed will, it is hoped represent “good” data.

This study has programmatic significance because it extends the research on task systems that began with Alexander (1982) and Tousignant (1982), and which generated a research program that has examined instructional conditions in physical
education using the ecological framework first proposed by Doyle (1977). Previous studies have had a descriptive focus. The current study has an experimental focus. Experimentally analyzing rate of response within the ecological framework of Doyle may extend the search for alterable variables that make a difference in physical education. These variables may in turn provide components for an instructional strategy that addresses Bloom's "2 sigma problem" in physical education.

Anthropological Assumptions

Siedentop (1983) proposed that research studies should include a brief statement relating to the view of humanity from which the study's methodologies are derived.

Such a section would not only alert the reader to the basic point of view of the researcher but, more importantly would require the researcher to consider seriously the implications of the questions asked and the assumptions underlying the implications of the questions asked and the methodologies used to answer those questions (p. 11).

This study was framed and conducted within the boundaries of the natural science of behavior. The subject matter of this science is the behavior of living organisms. Experimental analysis provides the systematic context for research in the natural science of behavior where behavior is the dependent variable in a system of functional relations that describe its lawful determination by classes of independent
variables. These independent variables include not only the present environment of the organism, but its ontogenic history, physiology, and genetic endowment (Pennypacker, 1992). Behavior is not an indirect means of studying something else such as cognition, will, or mind. These labels have behavioral referents that may be explained in terms of the effects of independent variables on behavior.

The two term contingency relation--the operant--is the basic unit of analysis in the natural science of behavior (Vargas & Vargas, 1991). An operant is usually described in its context (i.e., its immediate preceding stimuli) and the resultant is called a three term contingency consisting of a context or antecedent events, the behavior emitted, and the consequence/s. Operant relations which have occurred in the history of an individual continue to influence the probability of behavior which is yet to be emitted. Further, just as changes in the luminescence of the light in an operant chamber acquires control over the behavior of rat, variables in the current environment which are not necessarily the primary independent variables can also acquire control over the behavior of an individual.

These contextual variables have been labeled "setting factors" Kantor (1946), "setting events" (Bijou & Baer, 1978), contextual determinants (Gewirtz 1972; Morris, 1988) and establishing operations (Michael, 1982, Mallot, Whaley & Malott, 1993). The variables include, but are not limited to the immediate and observable discriminative stimuli in an
environment. Contextual variables include the past ontogenic history of the individual and the phylogenic history of the species.

The standard strategy in applied behavior analysis has been to hold contextual variables constant and to ignore their variation in order to search for a single, or a few controlling variables under the guise of a parsimonious explanation of events. But this approach has ignored (a) the fact that behavior is a continuous phenomenon (Johnston & Pennypacker, 1980) and (b) Skinner's (1957) notion of the individual as a temporal locus. A temporal locus is a point in time where the past and present contingencies (including the species phylogenic history) interact to produce a behavior. Morris (1988) has suggested that the search for contextual variables should be the subject matter of behavior analysis and that explanation of the variability of these variables is a necessary condition for a natural science of behavior.

The natural science of behavior holds the following assumptions:

1. Behavior is a phenomenon that occurs only at the individual level and it is continuous in its nature (Johnston & Pennypacker 1980).

2. The task of the natural science of behavior is to explain variability in the behavior of living organisms in terms of the organisms' interaction (functional relation) with the environment.
3. The environment exerts two principal types of control over our behavior. Control by the immediate consequences of our actions and control by the immediately preceding eliciting stimuli. These behaviors are termed operant and respondent respectively.

4. Descriptive studies provide information only on events and their occurrence. They do not provide information on their functional relations (Alexander, 1982).

5. Functional relations once identified must meet the standards of reliability and generality through demonstrated reproducibility at the level of the single organism (Pennypacker, 1992).

6. Experimental control replaces statistical inference as the tactical basis for judging the scientific utility of obtained results (Johnston & Pennypacker 1980; Pennypacker, 1992).

7. The cumulative development of science provides the only final answer to the importance of any particular data (Sidman 1960).

8. Teaching is the arrangement of contingencies of reinforcement under which students learn. They learn without teaching in their natural environments, but teachers arrange special contingencies which expedite learning, hastening the appearance of behavior which would otherwise be acquired slowly or making sure of
the appearance of behavior which might otherwise never occur. (Skinner, 1968, pp. 64-65).

9. Teaching is the term for class of responses which are directed at changing student behavior. Specifically, teaching involves developing new student behaviors, and maintaining or changing existing student behaviors.

10. Teaching effectiveness is defined by behavior change in the student.

Limitations of the Study

1. The study was limited to the observation of two experienced high school physical education teachers, in the same school and four students in each class. Data were collected for 30 and 31 lessons per class respectively.

2. The study was limited to the response classes of both the dependent and the independent variables defined in Chapter 3.

Definition of Terms

Accountability: "Keeping an account of the behavior of another to see whether it meets specification" (Skinner, 1974, p. 84)

Actual Task: The response that the teacher accepts in reference to the stated task.
Analog/s: “Something that is analogous or similar to something else” (Websters, 1990 p. 82). In this study the term is used to identify instructional strategies and problems in the gymnasium that are similar to those found in the classroom.

Arranged (contrived) contingencies: The preparation and application of the stimulus conditions (antecedent, consequent, or subsequent events) surrounding the responses of another person; producing contingencies different from those which would have occurred naturally.

Behavior: "That portion of an organism's interaction with it's environment that is characterized by detectable displacement in space through time of some part of the organism and that results in a measurable change in at least one aspect of the environment" (Johnston & Pennypacker, 1980).

Behavioral consequence: An environmental change (stimulus) that follows a given behavior in a relatively immediate temporal sequence and alters the probability of future occurrence of that behavior.

Behavioral outcomes: (probabilities of future occurrence)

a. The future rate of the behavior will increase.

b. The future rate of the behavior will decrease or
c. the consequence will either maintain existing conditions or have no effect on existing conditions.

Classroom ecology: The interrelationship (overlapping) between task systems (managerial, instructional, transitional and social)
Context: The current and past (historical) stimuli that discriminate an operant. Context or behavioral antecedents can be a singular stimulus as in a three-term contingency, or they can be nth-term.

Environmental change:

a). A new stimulus is presented or added to the environment.

b). or an already-present stimulus is terminated or removed from the environment.

Free operant: Any operant behavior where the opportunity to respond is not restricted, or inhibited. In physical education free operants are characterized by conditions where the teacher might phrase a task statement as “How many can you do” and where the opportunity to emit behaviors is not limited. Examples include dribbling a ball during basketball, or hitting a ball against a wall in tennis.

Intensity: A rate of student responding which is a function of teacher behavior, and/or successful outcomes (i.e., response products). Intensity is always relative to other parts of the lesson or to other similar lessons.

Observer Agreement: The percentage of agreement for how often two observers watching the same event, equipped with the same definitions of behavior, see it as occurring or not occurring at the same time.
**Natural consequences:** Consequences subsequent to performance and which are produced by the natural environment in which the performance takes place.

**Response:** An observable and classifiable instance of any class of behavioral events under investigation.

**Response class:** Varied forms of responses which are functionally equivalent.

**Restricted operant:** Any operant behavior where the opportunity to respond is restricted, or inhibited. In physical education restricted operands are characterized by conditions where the teacher controls the opportunity to emit behaviors, such as requiring a set number of repetitions or by having available a small number of balls for practice.

**Standard practice:** This term represents the existing instructional conditions of the class and constitute the current treatment (i.e., normal or standard practice). This condition also serves as a baseline against which to compare the effects of the treatments.

**Tasks:** A set of implicit or explicit instructions about what a person is expected to do successfully with a situation (Doyle 1981).

**Task congruence:** The extent to which task performance corresponds to task specifications.

**Task specification:** A performance requirement containing three elements: (a) situation, (b) performance and (c) criterion (Mager, 1975).
Task systems: The process by which one or more tasks are stated, executed, and maintained, in dealing with a particular content (e.g., managerial, transitional or instructional). (Doyle 1979; Alexander, 1982; Marks, 1988)

Units of analysis: Refer to the specification of responses in terms of functional and objective descriptions of behavior (Johnston & Pennypacker, 1980).

Units of measurement: Refer to standard and absolute methods or quantification such as the frequency with which a teacher speaks to a particular child; or the amount of time (duration) that a student spends engaged with the task (Johnston & Pennypacker, 1980).
CHAPTER II
REVIEW OF LITERATURE

Student responding has been used as a dependent variable in research on teaching for over three decades. A variety of labels have been used to describe responding (e.g., Academic learning time, time on task, success and error rates, percentage correct/incorrect). Responses have been measured using various tactics including number, rate, percentage and time (e.g., duration and latency).

Rate of response was the dependent variable used in this investigation to examine the effects of three interventions which are best described as instructional arrangements. The investigation tested the validity of the claim that rate of response is an alterable variable in instructional effectiveness research (including teacher-led as well as other forms of instruction) and describes the variability in the rate of response as a function of the different interventions.

Accordingly, this chapter reviews the educational research literature and synthesizes the empirical evidence that describes student responding and the instructional arrangements that affect student responding. The purpose is to
situate the present investigation within the research on teaching literature. The review will then examine the empirical literature in physical education which has reported on the level, nature and variation in responding in descriptive, correlational and experimental studies. Since rate of response is central to this review of literature, a definition and a brief discussion of rate of response is warranted before beginning the review.

Rate of Response

Johnston and Pennypacker (1980) define rate of response as behavioral events which occur in a unit of time (e.g., number per minute). In this definition the unit of analysis is a response (or a response class) and the unit of measurement is rate (or frequency). Rate of response can also be represented relative to the time in which it occurred. This measure, the rate of change in the rate of responding, is called celeration (Lindsley, 1991b).

Rate is commonly used synonymously with frequency. Though similar, rate and frequency differ in regard to the measurement periods in which a number count is made. The latter refers to measures in which the periods of time have been held constant (i.e., all observations occurred for 1 minute) and the former refers to responses which occur in intervals of varied durations, which are converted to a standard and therefore comparable (e.g., rate per minute).
Epstein (1982) summarized the advantages of rate of response as a dependent variable in measuring learning:

It reveals orderly moment-to-moment changes in behavior, is sensitive to change over a great range, and corresponds to what we usually mean by learning, as well as to the concept of probability of response. Rate of responding is affected by variables such as level of motivation and difficulty of response (Epstein, 1982, p. 41.).

Rate of response has been used extensively as the primary datum in basic (as opposed to applied) experimental settings where motivational variables (e.g., establishing operations such as states of food or water deprivation) have been used as independent variables to investigate the principles of behavior (e.g., reinforcement, punishment & extinction) on non-human animals. “By using rate of responding as a dependent variable, it has been possible to formulate the interaction between an organism and its environment more quickly” (Skinner, 1969, p. 7). Basic research scientists in behavior analysis have used rate of response as a datum for the past 50 years. Applied researchers in fields such as education and the helping professions (e.g., psychology, & social services) have increasingly used it as a dependent variable in their research.
Response as a Unit of Analysis in Education

The process-product era of educational research produced relationships between scores on achievement tests, and teacher and student behavior (Shavelson, Webb, & Burstein, 1986). Measures of teacher effectiveness were based on gains in student achievement. Correlational studies (e.g., Anderson, 1976; Brophy, 1973; Kounin, 1970; Lahaderne, 1968; Stallings & Kaskowitz, 1974) served to identify variables used in instructional packages that were then used in experimental studies (e.g., Anderson, Everstson & Brophy, 1979, Becker, 1977; 1982; Emmer, Everstson, Sanford & Clements, 1982; Emmer, Everstson, Sanford, Clements, & Worsham, 1982; Fitzpatrick 1981, 1982; Good & Grouws, 1979).

The search for explanatory mechanisms for the efficacy of daily instructional practices (particularly the behaviors of teachers and students) in terms of achievement gains led to the mediating-process paradigm. The mediating-process paradigm postulated time spent engaged with curricular content as the link between teacher process variables and product variables (achievement test scores). A major conclusion of this line of inquiry was that pupils who spend more time actively engaged with the task at a high success rate learn more of the content (Berliner, 1979). This conclusion was foreshadowed by the earlier work of Carroll (1963) and Bloom (1976).

How teachers arranged lessons in terms of time (instruction versus management time) and how students spend
their time (on task versus off task) have become important mediating variables in understanding teaching effectiveness defined in terms of student achievement.

Of the variables used to describe time in terms of student behavior the "academic learning time" variable (ALT) developed and used in the Beginning Teacher Evaluation Study (BTES) has demonstrated the most utility (Berliner, 1979). Academic learning time refers to that portion of the lesson when the pupil is involved with the criterion materials at a success rate of generally 80% or more (Fisher et al., 1980). Academic learning time differs from engaged time or time on task because it adds another dimension, that of success (or error) rate. Reviews by Berliner (1979), Brophy and Good (1986) and Smyth (1981) attest to the utility, efficacy and generalizability of the ALT variable.

The mediating-process paradigm used time not only as its unit of measurement, but frequently as its unit of analysis. The amount, magnitude and rate of behavior were hidden under a myriad of descriptors that referred to teacher response classes (e.g., teacher monitoring, active supervision, teacher prompts) and/or student response classes (on task behavior, student engagement, academic learning time) that occurred during a period of time. In short, the duration of the occurrence was recorded, but not the frequency. Often this was compounded by the use of systematic observation instruments that used interval recording (partial, momentary and whole interval)
tactics that obscured the differentiation of separate responses (a response occurring in one interval could not be discriminated as the same or a new response occurring in another interval).

Academic learning time added two previously neglected variables, success/error rate, and task difficulty, to time engaged. This was the first time that qualitative judgments about a response (i.e., correct versus incorrect) were used. But the actual success and error rates were still masked by the use of the interval recording tactics and by the nature of the ALT variable (i.e., time based decisions).

Much of the rationale for the mediating process paradigm was derived from Carroll (1963), whose model of school learning posited an inverse relationship between time needed to learn and achievement. In general if instructional time is held constant for all students in a group, achievement will vary as a function of ability. Carroll (1963) suggested that if students in a class received the same instruction in a subject, and the same amount of time to learn it, the resulting scores on achievement tests would be normally distributed. He proposed that achievement be held constant and time be varied, an instructional form of “to each according to his/her needs.” Students who had the least achievement would receive the most instruction and as they achieved more they would receive less instruction. The effect of this rearrangement of time and achievement would produce scores on an achievement test that were not normally distributed, but were instead highly skewed
with minimal variance (i.e., all of the scores would lie around the upper percentiles of an achievement test).

Carroll’s dependent variable was “the amount learned”. Five constructs were posited as independent variables. Three were designated characteristics of the student, ability (an individual's style of learning), aptitude (the amount of time the student will need to learn the task), and perseverance (the amount of time the student spends learning the task). Two were considered characteristics of instruction, opportunity to learn (the amount of time the teacher spends learning the task) and quality of instruction (the degree of instructional efficiency relative to the time needed for student achievement). These independent variables represent large response classes that are not well defined or delimited. For example both Carroll (1963) and later Shulman (1986) noted that the variable “quality of instruction” is difficult to define precisely and it “remains frustratingly elusive” (Shulman, 1986, p. 15).

Carroll’s (1963) work is important because it provides an intellectual framework using time as a unit of measurement and behaviors that occur within time as the unit of analysis for lines of inquiry such as like the Beginning Teacher Evaluation Study (Fisher et al., 1978), Bloom’s (1984) Mastery learning model, Rosenshine’s (1979) direct instruction and Good’s Active Teaching (1982) constructs.
Empirical Evidence of Responding in Education.

Good and Grouws (1977) in their study of fourth grade math instruction, administered the Iowa Test of Basic Skills to 43 classes. From this group they selected nine high scoring classes and nine low scoring classes (N = 18). They collected process and product data on all 18 teachers. One of the process variables was called rate of correct answers (actually a percentage of correct answers). Good and Grouws found that the rate of correct answers distinguished between more effective and less effective teachers. More effective teachers had a mean correct answer rate of 83% while less effective teachers had a mean of 76%.

One of the major findings of the BTES study was that a high percentage of correct answers was positively correlated with achievement (Fisher et al., 1978). This provided the rationale for the inclusion of a criterion of success in the ALT variable. Findings relating to the percentage of correct answers varied by context and activities. This suggests that perhaps different rates are necessary for different activities. In general Fisher et al., (1978) concluded that high achievement was associated with medium to high percentage of correct answers to questions posed by the teacher with the 65-75% (correct) representing a high percentage and 25-35% (correct) representing a low percentage.

In their study of first grade readers Anderson et al, (1979) found that the percentage of correct answers was positively
correlated \( (r = .49) \) with achievement gain. The study reproduced a similar finding by Fisher et al., (1978). Anderson et al, (1979) compared treatment (N=17) and control (N=10) classes over a half of a school year. Systematic observations were made once weekly in 10 of the treatment classes (7 treatment classes remained unobserved to test for observer reactivity). The study investigated the use of a small group instruction package consisting of 22 principles for organizing, managing, and instructing small groups. Anderson et al, (1979) found the mean percentage of correct answers in the treatment group classrooms was 73%, but only 66% in the control group classrooms. The finding was that the small group instruction package produced a higher percentage of correct answers than did conventional whole-class instruction.

Gerstein, Carnine & Williams (1981) like Good and Grouws (1977) found that more and less effective teachers (determined by the achievement gains by students) could be distinguished by the rate of correct responses, which they called an accuracy rate. Using the DISTAR reading program Gerstein et al., (1981) reported that in more effective classrooms accuracy rates were near 90% while in the less effective classrooms the accuracy rates were less than 75%.

These studies (Anderson et al., 1979; Fisher et al., 1978; Gerstein, et al., 1981; Good & Grouws, 1977) were all conducted in elementary schools. A middle school study by Evertson, Anderson, Anderson and Brophy (1980) failed to produce similar
findings. There is scant research in the teacher effectiveness literature conducted in middle and high school settings. Consequently both the level and nature of student responding in middle and high schools remain a neglected area of research on teaching in education.

The elementary studies by Anderson et al., 1979; Fisher et al., 1978; Gerstein, et al., 1981; Good and Grouws, 1977 frequently defined responding in terms of students correctly answering questions or problems posed by the teacher (in reading or math). Brophy and Good (1986) in their synthesis of these and other process-product findings concluded the following with respect to questions and responding "Questions must usually (about 75% of the time) yield correct answers and seldom yield no response at all, and their [student] seatwork must be completed with 90-100% success by most students" (p. 360). This conclusion suggests that student responding a) should be successful (i.e., a low error rate), b) is related to opportunities provided by the teacher (in this case questions posed), and c) that responding during seatwork should be more successful than responding during teacher-led instruction. This last implication may be an artifact of the task relative to student knowledge of the content. For example, it may be that during teacher led instruction the task is to provide (teach) the knowledge, and that during seatwork the task may be to apply the knowledge. If true this would suggest that different tasks are suitable for different stages on learning (e.g., recitation for
learning facts versus independent practice for applying facts to problems).

In their review of "Teaching functions" Rosenshine and Stevens (1986) commented:

the percentage of correct responses also play a role in successful learning. The importance of a high percentage of rapid ("automatic"), correct responses is a relatively new idea . . . Although there are no scientific guidelines as to exactly what percentage of correct answers should be, a reasonable recommendation at the present time (suggested by Brophy, 1980) is an 80% success rate when practicing new material. When reviewing, the success rates should be very high, perhaps 95% and student responses should be rapid, smooth and confident (p. 383).

A high percentage of successful responding implies a low percentage of errors. Neither Brophy and Good (1986) nor Rosenshine and Stevens (1986) suggested or implied any answers to the question "How much correct responding is sufficient to teach the correct response?" While there is little empirical evidence in support of "a level of correct responding" for optimum learning, their does appear to be unanimous agreement on the level of error rates. As Rosenshine and Stevens (1986) noted "In summary, whether one uses hints, prompts, or reteaching the material, the important thing is that errors should not go uncorrected" (p. 386). Error analysis the important thing is that (recognition of different types of errors) and error correction
(what to do when a student makes a particular type of error) are important lines of inquiry for educational researchers to pursue. Little has been done in the way of error research in regular education, but recently a number of studies in special education (Barbetta, Heward, & Bradley, 1993; Espin & Deno. 1989; Singh & Singh 1985; Van Houten & Rolider, 1989) have began to produce the beginnings of an error correction technology that is being systematically replicated with different content areas and different settings.

Fluency. Among the variables not yet investigated in the regular education literature is the amount of responding that is necessary (and sufficient) to ensure achievement. Percent of responding is not a good unit of measurement since it reports only on the correctness of the response relative to the total response and hides the actual level of the responding (number or number per minute). Frequently adjectives such as rapid or intense are used to describe the magnitude of responding (as distinct from the typography or the correctness of responses).

One measurement system that has been used to help teachers (and students) keep a record of changes in responding is “Precision Teaching” (Lindsley, 1991b). These records (data) provide the stimuli for curriculum decisions such as changing the content, changing the method of instruction and providing evidence of achievement gains. Since the measurement periods are held constant, rates are typically referred to as frequencies.
Frequencies are presented as relative measures or celeration on a standard celeration chart (a semilogarithmic graph with the Y axis ruled logarithmically and the X axis ruled with equal arithmetic spacing). As Lindsley (1991b) explains:

Precision Teaching took the slope (number per minute) of Skinner's cumulative records and charted it up the left of the standard celeration chart on a logarithmic scale. Calendar days were charted across the bottom, taking 140 days (20 weeks) to cover one school semester. . . . The frequencies up the left of the chart went from 1 per day through 1 per minute up to 1000 per minute. This covered six times ten cycles (log cycles) on the chart. The size was adjusted so that a line from the lower left corner to the upper right corner (an angle of 34 degrees) represented a doubling of the frequency every seven days. This is a celeration of times 2 per week. (p. 256-257)

In general Precision teachers follow the rule that if the celeration is below times 2 per week it is time to change something (content or the form of the instruction). Since these decisions are made on the basis of the learner's own behavior (presented on the standard celeration chart) the maxim "the learner knows best" commonly used by Precision teachers reflects that the basis for curriculum decisions is the learner's immediate past behavior.

Precision teachers/researchers using frequencies plotted on the standard celeration chart have described mastery as a
product of skills built through fluency (Johnson & Layng, 1992). Fluency is a measure of the accuracy and the rate of the response. Fluency is defined by Johnson and Layng (1992) as "the rate of performance that makes skills not only useful in everyday affairs, but also remembered even after a significant period of no practice" (p. 1476).

Haughton (1972, 1980) developed a set of fluency aims for common classroom content such as math and reading. A fluency aim is not a ceiling measure as in 100% correct, but provides for the teacher an aim or goal to plan for during instruction. For example, if recalling facts about a story read is the behavior of interest, a comprehension fluency-aim might be 20 facts recalled in one minute. The significance of fluency aims is that they require high frequencies of correct responding and indicate a level of responding that is sufficient for mastery (defined in terms of fluency) of the content.

White (1988) reviewed several experimental studies which have validated Precision Teaching in a variety of settings including elementary and high schools using regular and special education students. After his review of the major studies in Precision Teaching, White (1988) concluded: "In sum, Precision Teaching appears to be an effective, cost efficient, set of procedures for evaluating and improving instruction" (p. 530-532)

Fluency in context. Fluency is not an end in itself, it is part of an overall skill development strategy. This strategy has
been termed "generative instruction" by Johnson (1990, 1991) and Johnson and Layng, (1992). Generative instruction begins first with the assumption that one of the teacher's first instructional tasks is to establish (and build to fluency) the key component skills necessary for a particular behavior to occur. For example if the task is to answer math problems on a work sheet at a fluency of 50 per minute, then an important consideration is "Can the child write the appropriate numerals that fast?" Generative instruction would focus on this underlying skill of writing numerals fast (probably simultaneously with the math instruction) in order to develop fluency in answering the math problems. Johnson and Layng (1992) proposed a four step model of generative instruction. Figure 1 presents "The Morningside Model of Generative Instruction."

Students first establish accuracy, then build frequency of accurate responding (fluency) to the specific aim (fluency-building). This fluency is maintained for increasingly longer periods of time (enduring) and then applied (generalized) to new settings, activities or skills. Student's progress is measured using the standard celeration chart and progress is defined as a times 2 (or doubling) of frequency of change per week.
Figure 1. The Morningside Model of Generative Instruction

Note: The large arrows indicate that steps can move horizontally to provide overlapping learning phases. True mastery is the product of progressing up the steps.

(Adapted from Johnson & Layng, 1992, p. 1477)
Instructional Arrangements Influencing Student Responding

Carroll (1963) and Shulman (1986) acknowledged that the most abstract variable used in Carroll's (1963) model of school learning was "quality of instruction." The process-product and mediating process paradigms have been limited in their insight into the "quality of instruction," because they viewed the teaching learning process from the position of teachers' effects on students called "teacher effects" (Brophy & Good, 1986). Alternatively Doyle (1979) used the term "student effects" to indicate that students affect what teachers' do. This dual directional view (in fact multidirectional would be more correct, when one considers teachers often number 1, while students are many) has become a fundamental tenant of the ecological paradigm. The term ecological refers to a classroom ecology or "a pattern of relations between organisms and their environment" (Websters, 1990, p. 395). The ecological paradigm views classrooms as a complex and changing environment. In this view student responding is seen as influenced by "different segments and configurations of classroom activity" (Doyle, 1979, p. 140) mediated by teacher and student interaction.

The arrangement of the learning conditions in the classroom is thus a mediating variable in the level of skill involvement by students. For example studies by Anderson (1976), Fisher et al., (1978), Gerstein et al., (1981), Good and Grouws (1977), Gump (1969), and Kounin and Gump (1974) distinguish between some measure of time on task (or rates of
correct responding) obtained during seatwork, recitation, independent practice, and small group activities.

In addition to the finding that activities and class grouping influenced level of responding, there were also instructional arrangements that produced either accurate responding (low error rates) and/or high levels of overall (or total) responding. These instructional arrangements include Precision Teaching (Lindsley, 1990, 1992; White, 1986), (Anderson & Burns, 1987; Bloom, 1984; Guskey & Gates, 1988; Kulik, Kulik, & Bangert-Downs, 1990); tutoring (Bloom, 1984; Cohen, 1986); classwide peer tutoring (Greenwood, Delquadri, & Carta, 1988, Greenwood, Delquadri, & Hall, 1989; Greenwood, Terry, Arreaga-Mayer, & Finney, 1992); instructional alignment (Cohen, 1987); personalized systems of instruction, (Keller & Sherman, 1974; Buskist, Cush, & DeGrandpre 1991; Sherman, 1992); direct instruction (Rosenshine, 1979); Direct Instruction (Becker & Carnine, 1980, Kinder & Carnine, 1991); and programmed instruction (Skinner, 1968, 1989, Vargas, & Vargas, 1991).

These instructional arrangements have been labeled "mastery learning models" by Becker (1986) who reports nine features commonly associated with mastery learning models:

1. Objectives are specified, along with subobjectives grouped in small units for instruction.
2. Preskills are tested to ensure that placement is appropriate.
3. Procedures are developed to motivate and engage the student in active learning. These include reinforcement for working hard, for improvement, and for success.

4. Attempts are made to develop "quality instruction" that teaches the targeted objectives effectively and efficiently.

5. Differential time is allowed for different students to reach mastery.

6. Ungraded frequent testing is provided to monitor progress during early learning and provide feedback to the student.

7. Corrective-remedial procedures are provided if an approach fails.

8. Adequate practice to mastery of subskills is provided.

9. There is testing for longer term mastery of objectives.

(Becker, 1986, p. 156)

Mastery models teach skills to mastery in a cost efficient and a time efficient manner (Ellision, 1986). When they are compared to conventional instruction, they typically produce effects that are robust (changes which are of considerable magnitude and which endure) and educationally significant (i.e., students master the content to a significant criterion level).

Three additional features of the mastery models, tutoring, high levels of responding, and error correction are relevant to the current investigation. All of the mastery models present instruction in a tutoring style (using the teacher, a peer, a
computer, a programmed text). They all provide a high amount of responding which is often controlled by the speed of the students past response (another way of saying this is that the curriculum is presented at a rapid pace), and all remediate errors as they occur. Bloom (1984) argued that the solution to his “2 sigma problem” lay in the search for what he called “alterable variables.” These features of mastery learning (tutoring, high amounts of responding, and error remediation) are alterable variables in most if not all instructional settings and as such represent important independent and dependent variables in research on teaching.

A Summary of Responding in Education

1. Frequently phrases such as “rate of learning” or “engagement rates” (Bloom, 1981; Carroll, 1963; Fisher et al., 1978; Fisher et al., 1980; Smyth, 1981) have been used in the literature to represent the time spent on an activity by a student (these are variations of Carroll's' perseverance and/or aptitude variables depending on the focus: time spent on task or time spent till achievement). None of these uses of rate meet the definition introduced at the beginning of this chapter “behavioral events which occur in a unit of time (e.g., number per minute).” These other measures (mostly of time on task or a percentage of responding) though valid and reliably measured do not provide direct evidence as to either the total amount of responding or the amount of correct responding.
2. Successful responding and low error rates are desirable and differentiate between effective and ineffective teachers and the instructional arrangements that they use.

3. The amount of responding necessary for mastery appears to be referenced only in the literature of Precision Teaching and is referred to indirectly in the general research on teaching literature as a curriculum that is presented at a rapid pace.

4. The mastery learning models which are reported to produce the greatest gains in learning may do so because of an ecological relationship between the instructional arrangement (tutoring or small group instruction), the amount of responding (as fast as the student can respond or at least at a rapid pace) and the errors that the student makes (immediate remediation).

5. A possible conclusion regarding the absence of rate of response as a independent variable in the process-product and mediating process paradigms might be that it has been a hidden variable. In experimental studies when an intervention is introduced, changes in the dependent variable are either subscribed to the establishment of the intervention, or the removal of the baseline or control condition. A third explanation is possible. What if the intervention is a necessary, but not sufficient explanatory variable. That is, there is a third, unmeasured, and possibly undiscovered variable producing the change that is a component of the original treatment that is responsible for the change. Such might be the case with
relationships between time and achievement. Time (engaged
time/time on task) might be a necessary condition, but not
sufficient to explain achievement. Inaccurate descriptions of
the independent variable may threaten both the reliability and
the validity of the treatment. To avoid this threat, the response
class of interest might be further defined to include error rate,
success rate and total rate of responding.

The Responding Literature in Physical Education

Physical education researchers have vigorously adopted
the variable ALT in their own research on teaching. They have
given it a physical education analog, academic learning time in
physical education (ALT-PE), and developed context-specific
systematic behavioral observation instruments to measure it
(Siedentop, Birdwell & Metzler, 1979; Siedentop, Tousignant &
Parker, 1982). There are also several instruments which have
been designed to measure versions of engaged time as well as
other process variables that have developed along side ALT-PE
(Darst, Mancini & Zakrajsek, 1983; Darst, Zakrajsek & Mancini,
1989).

Academic learning time in physical education has been
actively used as an indirect measure of pupil learning in
descriptive, correlational, and experimental studies in a variety
of physical education settings (Lee and Poto, 1988, Metzler,
1989). Data from these studies have answered questions such
as “How do pupils spend their time in classes?” (descriptive
studies); "What teaching behaviors influence pupil behaviors?" (correlational studies); and "Can teaching behaviors that increase student achievement be taught to teachers?" (experimental studies).

Data obtained from descriptive studies (e.g., Godbout, Brunelle, & Tousignant, 1983; Shute, Dodds, Placek, Rife, & Silverman, 1982) indicate that pupils do not spend much of the lesson time engaged in skill practice. Instead, students spend much of the time waiting for an opportunity to respond, or waiting for the teacher to organize the next activity, both of which are non-content related activities. Data obtained from experimental studies indicate that it is relatively easy to change student behavior and produce higher amounts of time where students are engaged in skill practice (e.g., Birdwell, 1980; Ratliffe, 1986). Studies using the ALT-PE variable have also revealed that the ALT-PE percentages are influenced by grade level (Godbout, et al., 1983) and the nature of the content being taught. For example in a study that compared ALT-PE in different physical education settings Metzler (1979) found low percentages in gymnastics, but percentages in volleyball and football received high percentages. Studies have also found that boys and girls are approximately equal in ALT-PE (Shute, et al., 1982), but that pupils with differing skill levels vary in their ALT-PE (Shute, et al., 1982).
This review of the physical education literature on responding will limit itself to:

1. Studies which have reported on the rate of response (usually in terms of correct, incorrect & total) in middle or high school settings (and excludes elementary school settings).

2. Studies which have collected data “in situ” which excludes the considerable “testing” literature on response based student performance measures such as pre-post test measures on skills tests, but includes any measures that were collected as part of the day to day responses made by students (e.g., OTR’s compared to a criterion, often called a trial).

Rate of Response as a Unit of Analysis in Physical Education

Descriptive Studies

Alexander (1982). Alexander used the three term contingency model (antecedent-behavior-consequence) of behavior analysis to describe one student’s responses during 11 lessons of a 26 lesson golf unit. Alexander’s study remains one of the most rigorous attempts to describe the context and nature of responses. His units of analysis included teacher responses (e.g., feedback, demonstrations) and student responses (e.g., student performance of a task). The relationship between teacher behavior and student behavior was examined using the temporal dimensions (latency, duration, rate, celeration and interresponse time) of each response relative to
the other, and as stimulus or consequent conditions acting as part of an ongoing chain of events.

Alexander identified two accountability systems that influenced responses. Under conditions where the teacher's focus was on management, the managerial system operated. The managerial system was characterized by clearly stated tasks ("line up for roll call on the line"), it was objective (a student was either present or absent during roll call), it was formal (a loss of grade/points if absent was recorded in the roll book), and it had reliable and probable features (it was a sure bet that if a student was absent points would be deducted).

Under conditions where the teacher's focus was on instruction the tasks were specified less precisely. Usually a performance criterion was unstated (and possibly not expected), in contrast to the managerial system the judgments made by the teacher during instruction relative to the performance of the students were often subjective ("hit under the ball a little more than you have been"), and consequences (both reinforcing and punishing) were not reliably administered (there is a logistical problem in a class of 1 teacher to 20 students that prevents the teacher providing feedback to 20 simultaneous responses at once) and these consequences other than in skills tests, are not formal (i.e., no point or grade loss if skills are poorly performed).
In Alexander's study response rates varied according to the tasks (putting versus irons) and the location for the practice of the task (e.g., practice net, football field, wrestling room). The mean response rates for the student over the 14 lessons was 2.7 per minute. Rates ranged from 0.4 per minute when the student was in a group of four hitting golf balls using woods, to 3.8 per minute during putting skills test. The skills test produced significantly higher rates of responding due to the nature of the accountability system. Under this condition the instructional contingencies more closely resembled those of the managerial system. The tasks were clearly specified, decisions were less subjective (are the hands gripping the putter correctly), at times objective (did the ball go in the hole, number of putts made), reliable (records were kept and the teacher directly observed the performance) and formal (student performance influenced points earned and hence the grade for physical education).

Alexander's study was the first to describe responding in physical education from a behavior analytic perspective. It was also among the first to examine responses in physical education "in situ" using an ecological perspective (though Tousignant, 1982 was the first to investigate the ecological nature of task systems in physical education by extending Doyle's 1977, 1979, work using a qualitative paradigm).
Graham (1986). Using a social interaction paradigm, Graham investigated student involvement in an eighth grade physical education class learning volleyball. Her purpose was to identify factors "that support and constrain students' academic performance (i.e., motor skill performance) within and across the 14 lessons of an instructional unit" (Graham, 1986, p. 16). Among the measures she used to report on motor skill performance was rate of response (indicating total amount of responding in 1 minute) and successful rate of response (indicating the amount of successful responding in 1 minute). Six students were randomly selected from a teacher-produced list of high (from which 3 students selected) and low (3 students selected) skilled students. Data were presented graphically and visual comparisons were made between the total amount of responding and the amount of successful responding between the two groups of students. Data on responding was collected for 5 days (lessons 1, 2, 8, 12 & 14). Figure 2 presents a graphical display of student responding reported in Graham's (1986) study. In general, the data trends for target students were remarkably similar. There was however, considerable variability in the level of responding from the first lesson to the last. Rates for total responding began at 10 or 12 responses per minute and descended to 4 or less by day 14. Rates for correct responding followed a similar, but lower trend.
Figure 2. Rate per minute of total responses and successful responses for high skilled and low skilled students in Graham's (1986) study. Circles represent total responses and triangles represent successful responses. Brenda, Jody, and Dustin are high skilled students. Criag, Fred, and Tina are low skilled students.

Neither skill level, nor the gender of the student served to discriminate student responding. Graham (1986) suggested that differential student responding was a function of the tasks that the students were required to perform. When the tasks were
easy responding was high, when the complexity of the task increased responding decreased. Students then had less opportunity to master the skill than they had previously. She reports a range of 9-13 responses per minute during easy tasks (e.g., self passing) and a range of 2-10 responses per minute during more complex tasks. When Graham compared the successful responses with total responses for each of her target students the trend of more to less responding remained from day 1 to day 5. But, data paths of successful student responding were different. The difference (i.e., errors) between the total amount of responses made and the number of those that were successful was least for the high skilled students than it was for the low skilled students. This indicates that though both groups had relatively equal opportunity to respond, the amount of successful responding was unequal. In short, what was easy for the high skilled students was not easy for the low skilled students. Graham commented on this finding, "Not only were the high skilled students consistently successful more often than low skilled students during the lesson, the higher skilled students were considerably more successful." (p. 342).

Graham’s (1986) findings (relative to this rate of response) might be summarized as follows:

1. Students appeared to receive equal opportunity to respond without regard to ability or gender.
2. The variable successful responding differentiated between ability levels, but not gender (though gender was not a focus of the study she did indicate the respective gender of the students and thus makes this conclusion possible). High skilled students were more successful more often than low skilled students.

3. Task complexity (e.g., tasks such as drills involving 3-4 people interacting with each other, though the task itself might be quite easy) influenced the overall level and success of responding. To determine the influence of task complexity Graham examined the nature of the task. She found that when students (both high and low skilled) practiced tasks that were highly complex the level of total responding was similar, but the level of successful responding differentiated between the two groups.

Methodological Comments

1. Graham used a sample of 5 of 14 lessons (specifically lessons 1, 2, 8, 12 & 14). Five days represents a small sample. Some caution is therefore warranted in interpreting the results.

2. Rate of response was calculated for all responses made during the engaged time of the lesson, then changed to a rate per minute. This rate is a mean rate for the lesson. We do not know the actual rate per minute for each student. Thus we cannot determine if rate varies throughout the lesson as a function of
different tasks and/or time (i.e. when in the lesson a particular rate occurred).

3. Rate of response was calculated across tasks. At times the tasks varied across students, as a consequence of student modification of the task and also because at times a choice of tasks was offered. Rates for different tasks may not be equal! If rate of response is affected by tasks (i.e., the context) then a useful strategy might be to represent rates relative to each task, in a similar way to the way percentages of ALT-PE are discussed relative to different contexts (e.g., volleyball versus gymnastics).

Son (1989). Son analyzed task congruence (between the stated and the observed task) in Korean middle school physical education classes. Task congruence is an important variable in the responding literature because “What the teacher presents as the task is not sufficient to explicate what students learn; what students actually do to meet the task demands is a better explanation of the student learning experience” (Son, 1989, p. 9). The tasks were examined using the three term contingency framework first used by Alexander (1982). Son used student responses as a dependent variable to describe how teacher specification of tasks (e.g., explicitly versus implicitly stated) influenced the number of responses (which he called frequency), the percentage of congruent responses and the nature of responses (e.g., successful, unsuccessful, correct, incorrect).
Son observed a class of 59 students for 12 weeks. Thirty-six lessons were observed. Each lesson lasted for 45 minutes. The relevant findings were:

1. Student responses differ (by number and percent of total congruent responses) across different sports. Son examined instructional units of fitness (455 total responses / 75% congruent responses), tumbling (183/89%), team handball (311/79%), and soccer (260/85%).

This finding validates ALT-PE (proxy measure) findings that indicted that responding differed across activity (Metzler, 1989). Though tumbling had the lowest total responding (a finding supported by Metzler, 1979; Ward, 1988; 1990) it had the highest percentage of congruent responses. Son reported on aversive control (threats & punishment) used by the teacher during the lessons on tumbling. One possible explanation for low total responding and a high percentage of congruent responses might be that students engaged in tumbling were under strong aversive control by the teacher. Since safety is an important consideration in tumbling exercises perhaps the aversive control was strongest here to prevent injury. Equally possible might be that the task difficulty was easiest here to prevent injury.

2. Son reported that the responding data differentiated between students of different ability levels (which he described as above and below average students). A finding that is also found in the ALT-PE literature (Shute, et al., 1982) and by the
earlier work in task systems by Graham (1986). Table 1 presents the mean percent of responding for high and low ability students in Son's study.

Table 1
Mean Percent Responding for High and Low Skilled Target Students Across Classes (Son, 1989)

<table>
<thead>
<tr>
<th>Ability Level</th>
<th>Correct/ Successful</th>
<th>Correct/ Unsuccessful</th>
<th>Incorrect/ Successful</th>
<th>Incorrect/ Unsuccessful</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>83.9%</td>
<td>6.3%</td>
<td>5.4%</td>
<td>4.4%</td>
</tr>
<tr>
<td>Low</td>
<td>22.8%</td>
<td>16%</td>
<td>30.7%</td>
<td>30.5%</td>
</tr>
</tbody>
</table>

Note. A correct response was a response that was topographically correct. A successful response was one in which the outcome of the response is successful.

The data were analyzed according to two criteria, correct and successful responses. A correct response was a response that was technically or topographically correct. A successful response is one in which the outcome of the response is considered successful (e.g., the ball goes through a hoop in basketball). A response could have elements of each (e.g., correct but not successful). This creates 4 possible
combinations of the criterions (see Table 1). The data in Table 1 for the high ability students indicate that these students were mostly correct and successful (84% of the time) with minimal error (5% or less) in each of the potential error categories (combinations of incorrect and unsuccessful). In contrast, students designated with low ability were correct and successful only 23% of the time and were both incorrect and unsuccessful for approximately 30% of their responses.

3. A third finding identified by Son sheds some light on this variability. Son suggests that the “student responses are more affected by the nature of the task itself than by task specification by the teacher.” In short, what is difficult for one student may not pose any difficulty for another and since ability is the most influential variable of task difficulty the findings presented in Table 1 might reflect a “one size fits all approach to instruction” and therefore the considerable variability is an artifact of task difficulty. This finding extends the research into task systems begun by Tousignant (1982), it is also analogous to the argument made by Carroll (1963) who argued that achievement varies as a function of ability.

Methodological Comments

1. Son used the term “frequency” throughout his study, but did not report on the different durations of each lesson. The frequency measure used is “number of responses per lesson.” Though the length of the lessons were 45 minutes, ALT-PE
studies have shown that the time available for activity and the engaged time (ALT-PE) differ across lessons. Responses per lesson are a crude indicator of responding throughout the lesson and responding to different tasks.

2. Son did present levels of responding across different types of tasks. He is among the first to do so. His findings suggest that refining tasks produced the highest amount of responding, and informing tasks the lowest. A caveat is warranted. This finding is based on the aggregate of all sport tasks (fitness, tumbling, team handball, & soccer) and as such does not reflect the differential responding to different tasks in each setting. None-the-less the tactic provides further insight into the differential nature of responding.

Lund (1990). Lund used rate of response as a dependent variable to indicate the effectiveness of different accountability systems used by 5 high school physical education teachers. Four target students (including both high and low skilled) were observed in each class. The teachers all taught volleyball, but were free to arrange the instruction as they saw fit (i.e., as they usually taught volleyball). All the teachers had coached, or were coaching volleyball.

Rate of response was described using response rate per lesson, and response rate per minute. In addition a percentage of correct responses (i.e., topography) and a percentage of
successful responses (i.e., outcomes) were also used to provide some qualitative measure of the response and its outcome.

Two of Lund's (1990) research questions are relevant to the current investigation.

1. How do accountability systems affect student response rates in terms of total number emitted, topographical correctness and success? Differential responding of high and low skilled students was also examined in answering this question.

2. Do student response rates vary among teachers?

Table 2 and Table 3 present the data obtained by Lund (1990) for mean response rates per minute (Table 2) and mean response rates per lesson (Table 3). Data are arranged across each of the 5 teachers and differentiated on the basis of more or less (i.e., high & low) skilled students.

Response rates within classes varied by skill level. The class with the most responding taught by Mrs. Brown, had high rates of responding for high and low skilled students in Table 2 and Table 3 (mean responses per minute and per lesson). In short, both high and low skilled students in her classes made more responses per class and in all lessons in the instructional unit, than did the students in the other classes. Mrs. Brown's low skilled students responded more than did the high skilled students in the other classes with the exception of Mr. Far.
Table 2
Mean Response Rate per Minute for More and Less Skilled Target Students Across Classes (Lund, 1990)

<table>
<thead>
<tr>
<th>Target Student</th>
<th>Adams</th>
<th>Brown</th>
<th>Camp</th>
<th>Dunn</th>
<th>Far</th>
</tr>
</thead>
<tbody>
<tr>
<td>More</td>
<td>2.06</td>
<td>3.65</td>
<td>1.65</td>
<td>1.57</td>
<td>2.04</td>
</tr>
<tr>
<td>Less</td>
<td>1.66</td>
<td>3.47</td>
<td>1.22</td>
<td>1.42</td>
<td>1.69</td>
</tr>
</tbody>
</table>

Table 3
Mean Response Rate Per Lesson For More and Less Skilled Target Students Across Classes (Lund, 1990)

<table>
<thead>
<tr>
<th>Target Student</th>
<th>Adams</th>
<th>Brown</th>
<th>Camp</th>
<th>Dunn</th>
<th>Far</th>
</tr>
</thead>
<tbody>
<tr>
<td>More</td>
<td>45.33</td>
<td>58.56</td>
<td>36.70</td>
<td>46.64</td>
<td>58.83</td>
</tr>
<tr>
<td>Less</td>
<td>36.44</td>
<td>54.44</td>
<td>28.70</td>
<td>38.86</td>
<td>48.83</td>
</tr>
</tbody>
</table>
Furthermore, Lund's qualitative data indicate that more of the responses made by Mrs. Brown's students were correct and successful, than were the responses made by students in the other classes (or as Lund would say students operating under different accountability systems produced different levels of responding). Lund attributed the differential responding across classes to the difference in the nature and application of the various accountability systems found in each class.

Surprisingly, Mrs. Brown had the lowest activity time of all the teachers. In part, this can be explained by the time involved in implementing Mrs. Brown's accountability system. Mrs. Brown's system took time to implement and this time was at the expense of time that might have been available for instruction. Nonetheless in Mrs. Brown's class more work was done in less time by students than in any of the other classes. Brown also had two to three times as many tasks per lesson than the other teachers (10 versus 3 or 4 tasks per lesson) suggesting that Brown allowed a short period of time for each task to be completed, but as evidenced by the higher rate per minute, more responding (which was also correct and successful) was accomplished in this short period of time. Mrs. Brown also had lowest response rates in games compared to the other teachers. Another teacher Mrs. Dunn had allocated the most practice time, and had highest amount of game time, but her student's uniformly produced the lowest rates of responding and the lowest percentages of correct and successful responding. In comparing
Mrs. Brown and Mrs. Dunn one might conclude that having more activity time is not sufficient to ensure high response rates (although activity time is clearly necessary). Lund noted that "the responses that a student makes indicate the real amount of practice that a student had in a lesson, rather than the activity time" (1990, p. 199). Lund's comment is a reflection that proxy measures of achievement (e.g., activity time, ALT-PE) do involve some inference and that the inference is not always accurate when compared to a direct measure in this case rate of response. Lund's statement "the responses that a student makes indicate the real amount of practice that a student had in a lesson, rather than the activity time" is also consistent with the assumption made by Son (1989) who based his study on the premise that "What the teacher presents as the task is not sufficient to explicate what students learn; what students actually do to meet the task demands is a better explanation of the student learning experience" (Son, 1989, p. 9).

Lund found that high skilled students tended to have more responses than low skilled students, and that these responses were usually correctly performed (technique) with successful outcomes (the ball went over the net). With regard to successful responding Lund concluded "rate of success is a function of skill level, and the difficulty of the curriculum as well as accountability" (1990, p. 189).
Comments:

1. Like Graham (1986) Lund's study provides some evidence that the nature of the task influences the level of responding. Though she did not correlate the rate of response relative to task type, Lund's particular contribution is the recognition that responding is also influenced by different types of accountability systems.

2. Lund used rate of response to quantify responding and percentage correct and successful to qualify the nature of the response. Lund's study continued the focus on qualitative dimensions of the response.

3. One important feature of Mrs. Brown's class (the most effective teacher in terms of the quantitative and qualitative responding) was that the class moved at a rapid pace including rapid presentations of tasks and prompting of on-task behavior by the teacher. Additionally the high response rates and the low variance between high and low skilled students in Mrs. Brown's class suggests that increasing the amount of responding may be an important (and alterable) variable in explaining successful performance in physical education.
Experimental Studies

Dugas (1983). Dugas investigated the relationships among ALT-PE, opportunities to respond (OTR), and the number of criterion trials (CT) as process variables and measures of student achievement (product variables) in high school archery classes taught by four student teachers. Opportunities to respond were operationalized by Dugas as the total number of responses made by the students. Each OTR was further analyzed in terms of its topography (acceptable, not acceptable or no response made when one should have been made). Criterion trials were used as an outcome based measure. Fifteen classes for each teacher were observed. Scores on the A.A.H.P.E.R.D. archery skill test (pre & post instruction) and a knowledge test were used to provide the product measures. Data were analyzed using the student and class as units of analysis. Pearson product-moment correlations (partial correlations) were used to identify correlations between process and product measures. In addition a one-way analysis of variance (ANOVA) was conducted to determine if interclass differences existed treating teacher and student process variables as independent variables.

Dugas found moderate to very high correlations between process variables using the student as the unit of analysis (OTR/CT r = .94; ALT-PE/OTR r = .43; ALT-PE/CT r = .56). Stronger correlations were found when the class was the unit of analysis (OTR/CT r = .99; ALT-PE/OTR r = .66; ALT-PE/CT r =
.70). Using pre and post test measures of student performance on the A.A.H.P.E.R.D. Archery test the most highly correlated process variable was criterion trials (using the student  \( r = .24 \); using the class as the unit of analysis  \( r = .99 \)). The ANOVA did not reveal any interclass differences in achievement scores that were a product of process variables.

The correlation between criterion trials and the skills test (a process-product correlation) lends further support to findings and comments made by descriptive researchers (e.g., Graham, 1986; Lund, 1990; & Son, 1989) who reported that student responses rather than time (duration) measures such as ALT-PE and engaged time are more accurate indicators of performance in physical activities.

Table 4 presents the mean number of response opportunities and the mean number of criterion trials per instructional unit per class for subjects in Dugas' study. The data reported in Table 4 are the mean data for the target students in each class. The data indicate that more opportunities and criterion trials were made by students at 10 yards than at 20 yards. From these data one can estimate the mean daily (lesson) responding at approximately 36 responses per student in this archery unit.
Table 4

Mean Number of Opportunities to Respond and Criterion Trials by classes in Dugas' (1983) study.

<table>
<thead>
<tr>
<th>Class</th>
<th>10 yds</th>
<th>20 yds</th>
<th>Total</th>
<th>10 yds</th>
<th>20 yds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150</td>
<td>72</td>
<td>222</td>
<td>140.7</td>
<td>70.7</td>
<td>211.4</td>
</tr>
<tr>
<td>2</td>
<td>168</td>
<td>84</td>
<td>252</td>
<td>168.0</td>
<td>83.3</td>
<td>251.3</td>
</tr>
<tr>
<td>3</td>
<td>139</td>
<td>54</td>
<td>193</td>
<td>135</td>
<td>46.4</td>
<td>181.4</td>
</tr>
<tr>
<td>4</td>
<td>198</td>
<td>42</td>
<td>240</td>
<td>196.7</td>
<td>41.3</td>
<td>238</td>
</tr>
</tbody>
</table>

Summary data in Table 4 indicate that with the exception of class 3 the total responses (including opportunities and trials) for students in the remaining classes are similar in number.

Comments

1. Dugas used mean/s (number of responses) as the unit of analysis (a requirement of the correlation analyses used) in her study. This data hides the nature of the performance over different distances. In particular partial correlations were performed on mean OTR's and CT's from distances of 10 and 20 yards. Students had 4.5 lessons devoted to shooting from 10 yards and 1.5 lessons devoted to shooting from 20 yards.
Achievement was lowest on the scores obtained from 20 yards. The analyses were conducted on mean data that were derived from periods of different length (i.e., 4.5 lessons versus 1.5 lessons). It is appropriate to ask if these analysis are “fair” analyses? Do the findings represent what would happen if the time for instruction for each distance were to remain equal?

2. The unit of measurement used to report the duration in which the the OTR’s and CT’s occurred was the lesson. However, lesson duration or engagement times were not reported (only number per lesson). If the available time for OTR’s varied across lessons, then variability in OTR’s and CT’s might also be a function of time.

3. Dugas also reported substantially different correlations using the class as the unit of analysis and the student as the unit of analysis. This was most evident when ALT-PE was correlated with the skills test. The correlation between ALT-PE (total intervals) and the skills test when the class was the unit of analysis was $r = .92$. However the correlation between ALT-PE and the skills test when the student was the unit of analysis was $r = 0$. When data are correlated by distance the correlation (for student as unit of analysis) is $r = -.17$. These findings raise some questions about the relationship between class and student as units of analysis. In short, what is the validity of the two sets of data? Which is the better indicator?
Pieron (1983). Pieron reported on the effectiveness of teaching a gymnastic routine using a 10 day (5 week twice per week lessons) instructional unit to one class of sixteen students. Eight students were observed (4 high skilled and 4 low skilled). Learning gains were measured by using an estimate of the errors in the performance. A pretest-posttest design was used and the data were analyzed using a Mann-Whitney U test (a t test for ranked data). No significant difference was found in student engagement time. However significant differences were found in favor of students learning more. Four variables were identified as significant: specific motor activity (p=.014), the number of trials (p = .029), the success rate (p = .014) and specific feedback (p = .171).

Pieron's research provides further confirmation

1. that response rate, in this case success rate and trials, is associated with achievement.

2. that student engagement is not sufficient in discriminating between different levels of achievement.

**Discrete Trials**

Several of the studies reviewed in this chapter have noted that response based measures (as opposed to time based process measures such as ALT-PE, engaged time) are more accurate (by their nature, they are not proxy measures) in describing the performance of the student. This issue was initially discussed by Siedentop, et al., (1982) who proposed and recommended the
use of discrete trials as measures of behavior change. Alexander (1983) made a persuasive argument in favor of discrete trials as measures of behavior change. Alexander described the problem:

In a gymnastics class there is no way to determine whether, of 20 MA [ALT-PE's motor appropriate category] intervals, all were forward rolls, 5 were forward rolls and 15 were flank vaults, or some combination of other skill responses was emitted” [parenthesis added] (p. 46).

Indeed if interval sampling techniques are used it is impossible to even quantify performance other than to report on the number of intervals that satisfied the criterion of 80% successful at an appropriate level of difficulty (see earlier discussion in “Response as a unit of analysis in education”).

Alexander argued that:

"Quantifying behavior only in terms of its duration does not guarantee the accuracy of behavior change. . . . Discrete trials procedures are capable of measuring behavior change directly while ALT-PE can only predict such changes. . . . Why be content with a predictor variable when more direct means of qualifying behavior change are possible (1983, p 45).

Alexander's comments are a call for direct measures of behavior change, not merely discrete trials. Discrete trials will not be an appropriate measure for all types of analysis"
(e.g., behaviors during game play), however direct measures should be sought for all types of analysis. There have been several studies in physical education which have investigated the relationship between discrete trials and achievement (Buck, Harrison, & Bryce, 1990; Silverman 1985, 1990).

Silverman (1985). Silverman investigated the relationship between two groups of process variables, student engagement and practice trials and achievement. Fifty-seven students (from an initial 102) enrolled in five intermediate swimming classes were taught 2 fifteen minute lessons each on the survival float. The students were pre and post test and the instructors were free to use the 15 minutes anyway they wished (i.e., the nature, type and amount of instruction was not controlled). This design has been commonly called an experimental teaching unit. An experimental teaching unit is a standardized unit of instructional content presented by teachers in a short instructional period. The ETU “package” consists of pre- and post-tests developed directly from the specific instructional content, a predetermined instructional time span, and stated instructional objectives. To reduce prior learning effects, the instructional task should be novel for the target subject group. Typically, teachers are free to choose their own instructional strategy (Metzler, 1983, p. 187).
On the basis of the pre-test students were categorized as low-, medium, or high-skilled. The skill tests consisted of 5 expert instructors rating the performance of the components of the survival float. Samples of ALT-PE for each student were also obtained. Silverman used multiple regression analysis to determine the relationship between the process variables and achievement.

The results indicated that

1. Appropriate practice trials and total trials are positive predictors of achievement in high skilled students.

2. Inappropriate practice trials are negative predictors of achievement in medium skilled students. Both of these findings (positive and negative predictors) lend support to the notion that more practice is better than less practice and more correct practice is best. In short, lots of perfect practice probably makes perfect.

3. No process variables used in this study predicted achievement for low skilled students. The wide variability in skill level of low and moderate groups may have been an artifact of the way in which the pretest was conducted. Silverman suggested initial repeated measurements may have given more data upon which to base a pre-test score.

4. No time engagement variable related to achievement when students were grouped together. Silverman concluded that “global measures of student engagement alone will not be the
most potent variable for predicting student achievement” (p. 19).

5. A variance of 26.4% (R²) in the post test scores was accounted for by the pretest scores. Silverman notes that this variance is much lower than similar classroom studies (referring to large scale process-product studies) and suggests this may be because of the greater variability in the initial skill levels in physical education than in the classroom studies.

Comments:

1. Experimental teaching units are a design arrangement where process measures are compared to product measures. In Silverman’s study the instruction lasted for two 15 minute lessons. Some questions regarding his design arrangement are appropriate. Is 15 minutes sufficient time to measure process variables? If the task (survival float) can be taught quickly (in 30 minutes) are total response-process measures a sufficient explanation for the change? For example consider a task that is relatively easy to perform where the primary process variable is a good demonstration. In such cases trying the task once or twice would produce mastery. A large response rate would be redundant, and would not function as an explanatory variable. The question is whether or not the survival float taught for 2x15 minutes lesson is a “valid” test for total response process measures (though qualitative measures such as
appropriate/inappropriate measures might be more valuable, or after how many trials was mastery achieved).

Silverman (1990). Silverman investigated linear and curvilinear relationships between student practice and achievement in 10 physical education classes conducted in middle and junior high schools. Four teachers each taught one class and 3 teachers each taught two classes.

Six skills were observed:
1. Serve without ball
2. Serve with ball but without net
3. Serve with ball and net
4. Forearm pass without ball
5. Forearm pass with ball
6. Forearm pass with ball against wall

The dependent variables were residual scores calculated for each student as the measure of achievement. The data were analyzed using the pearson product-moment correlation. The main effect was that there were fewer appropriate trials than there were inappropriate trials. Students spent more time getting it (the skills) wrong, than they did getting it right. The correlations for appropriate trials were .653 (p<.001) for the forearm pass and .654 (p<.001) for the serve (a positive and substantial association for both skills). The correlations for inappropriate trials were -.185 (p<.005) for the forearm pass (a
negative and low association) and -.615 (p<.001) for the serve (a negative and substantial association). The overall conclusion was that correct practice is positively related to achievement and incorrect practice is negatively related to achievement. This finding is consistent with the earlier conclusions of Dugas (1983), Pieron (1983), and Silverman (1985).

Buck, Harrison, & Bryce (1991). Buck, Harrison, and Bryce observed the learning trials for the volleyball set, forearm pass, serve and spike of 58 students in two coeducational beginning volleyball classes at Brigham Young University (students were undergraduates). The students were observed for 22 days and the resultant data were analyzed to determine the relationship between learning trials and three measures of volleyball achievement.

A percentage of correct trials per student (correct trials/total trials x100) was plotted using learning curves (the authors used a straight line linear model to explain performance change over time). Data were also graphically displayed by ability level (high, medium and low skilled groups) as determined by an initial skill pretest (A.A.H.P.E.R.D. Volleyball skill tests and the Stanley spike test). An analysis of covariance (ANCOVA) was conducted to determine the influence of the learning curve data, gain scores (posttest-pretest scores, adjusted for the pretest scores), and teacher ratings (scores
given by the teacher for skill performance during volleyball games) on each of the four volleyball skills.

Three principal conclusions were discussed by Buck et al:

1. "The most consistent result of the statistical analysis for all the achievement variables was the importance of the total correct trials in determining achievement" (p. 148). Buck et al., also examined performance outside-of-class (i.e., staying behind after class to play a few games). Not surprisingly, those students who participated in outside-of-class games had the highest level of correct trials. A reasonable explanation for the higher level of correct trials by students who stayed behind is that these students had more opportunities to practice. Buck et al., acknowledged the importance of correct practice, "Total correct trials was the only significant factor in determining the rate of learning for the set and the serve" (p. 148).

2. "Average trials per day per student were very low" (p. 148). Buck et al., report on a large variance in the trials from day to day. They recognized a general pattern of responding during the 22 days. On days when a skill was introduced (instructed) the most responding for that skill occurred. In the days following the introduction on the skill the skill was either not practiced again or practiced very little. For example Buck et al., noted

The average student attempted a set, forearm pass, serve, or spike 49 times per day during the 60 minutes available, contacting the ball less than one per minute. If the day
when the skills were taught were removed from the calculation, the values would be much lower (p. 141).

A response rate of "less than once per minute" is not a high response rate. Buck et al., also noted that very little performance improvement occurred from day 1 to day 22 for the students in these classes. Over 50% of the classes were devoted to game play. A review of these findings might be stated as follows. Students did little responding, some of the responding was correct, but the level of responding was insufficient to improve performance. Given these findings Buck et al., argued in favor of finding instructional methods that produce more correct trials per day and therefore improving performance. An unanswered question remains "How much responding is necessary to produce performance gains?".

3. If the average responding was low, it was lowest if you were a low-skilled student. Low-skilled students did not get as many trials as high skilled students in practice and especially during game play. Buck et al., commented "Whenever a significant difference between ability groups existed, the low-skilled ability group was always the group that differed" (p. 151).
Games Studies
Games are used to further develop skills in complex settings. Games are used to teach strategy. Games are used to provide experiences which are thought to produce beneficial outcomes in the affective domain. Games are thought to be ideal settings to learn about and practice skills related both to cooperation and to appropriate competition. And, lest we forget, games are usually a lot of fun for those who play them. (Siedentop, 1990, p. 70)

Siedentop (1990) continued by pointing out that descriptive-analytic research (mostly with a ALT-PE focus) has shown that games occupy a dominant position in the organization of physical education content second only to skill practice. Games can be analyzed in terms of their development of skill or in terms of their development of strategic play.

Few if any studies have focused on strategic play as most have used skill development as this focus (Siedentop, 1990). However there are several curriculum models which use strategic play as part of their foci such as “Sport Education” (Siedentop, 1985) and “Games for understanding” (Werner & Almond, 1990).

When skill development is the focus of games the student skill responses are the dependent variables and are analyzed the same way as they were in the responding literature previously
discussed in this chapter. Responses are defined quantitatively (the amount of responding) and qualitatively (typography/technique and outcome). The units of measurement have included number, percentage and rate. Descriptive-analytic as well as experimental studies have been used to investigate skill development in games.

The unit of analysis to date has been the individual. An individual when he or she is a member of a team provides is one link in a complex set of events which is called team performance. A team's performance can be and has been analyzed as chains of events (Eom & Schultz, 1992; Mace, Lalli, Shea, & Nevin, 1992; ). However a caveat is warranted. Mace et al (1992) commented that "it may be appropriate to view a sports team as an aggregate organism whose behavior is functionally similar to that of an individual subject" (p. 662). Treating the team as the unit of analysis. would be an erroneous tactic. Teams are a collection of individuals (much like cultures are a collection of individuals) and the analysis of the individuals performance should not be blurred nor ignored in favor of a team analysis, for without the individuals, there is no team.

Games as Special Conditions

When games are considered as a special condition (i.e., an independent variable) and responses are used as dependent variables the relevant questions center around the various
manipulations that can occur in the context called “the game.”
The variables that can be manipulated are extensive and include
student skill/ability level, gender, and teamsize. For example,
in a 3x3 volleyball game (volleyball is among the most studied
of sports) is the assumption that 33% of OTR’s to each player on
one team valid? Physical educators would expect OTR’s to be
evenly distributed in a setting such as volleyball. There are
questions about generalization from practice settings to
competitive settings such as “Are there special game-like
activities that provide better lead-ins to games?” and “Do
modifications for children's sport work?”

This section of the chapter examines descriptive and
experimental studies that use rate of response or responses as
their unit of analysis in game settings. The inclusion criteria
remains the same as for the responding literature previously
reviewed with two exceptions. The Tharp and Galimore (1976)
study is included because it represented a jumping off point for
inquiry into sport/games and responding. The Brown (1986)
study is included because though it dealt with upper grade
elementary school children it represents a line of inquiry that
has implications for studies of middle schools.
Tharp and Galimore (1976). Tharp and Galimore observed John Wooden during the 1974-75 NCAA college basketball season. They used systematic observation of his instructional behaviors to find out "How he talks to his team." Their findings are among the most cited in the sports pedagogy literature. Tharp and Galimore observed 2,326 teaching acts and categorized them into 10 categories. Five categories account for 85% of Wooden's coaching behaviors: Instructions (50.3%), Hustles (51.7%), Scold/reinstruct (8.0%), praises (6.6%) and scolds (6.9%). The purpose of these categories is to produce and maintain correct performance and reduce errors. Thus 85% of the content of Wooden's vocal behavior during practice was designed to improve or increase the quality or quantity of responding. These data and the anecdotal comments of Tharp and Galimore indicated that the instructional behaviors (including all of the above categories) were delivered at a rapid pace. Tharp and Galimore noted "He [Wooden] continually shouts instructions" and "Wooden values intensity" (p. 76). These comments and the descriptive data provided some indication of the pace of practice sessions and in particular the pace of player responding. While no data were collected on player responding Tharp's and Galimore's description of Wooden's practices suggested high levels of player responding.
This study is significant for at least four reasons:

1. The study is one of the first times that intensity of practice has been discussed (albeit indirectly) in the empirical literature. Sport-casters typically talk about an intense game or the intense level of performance though few researchers had discussed it prior to Tharp and Galimore.

2. The study was the first descriptive-analytic research on coaching in a college setting.

3. Wooden responded to individual player performance. That is Wooden’s coaching behavior was occasioned much of the time by individual player performance (principally unsuccessful skill performance).

4. The cookbook approach to the application of praise as reinforcement in instructional settings was questioned by Tharp and Galimore. Much of what Wooden did was to give instructions, and though there were near equal amounts of praise and scold/instruct behaviors, both were low (less than 7% each). Tharp and Galimore concluded that much of the reinforcement was provided by the environment other than Wooden (by other players, and through successful shots, plays etc.).

Mace, Lalli, Shea, & Nevin (1992). A phenomenon called “behavioral momentum” initially developed in special education (applied) and laboratory (experimental) settings was used to explain player behavior during college basketball games. Behavioral momentum is defined by two features (a) a rate of
response and (b) persistence. A similar, but different phenomenon called extinction occurs when reinforcement is withdrawn following a period of reinforcement for responding. In extinction the behavior will reduce or “undergo extinction” when the reinforcer is withheld. There is frequently an extinction burst (a brief increase in responding), but this is short lived in the absence of reinforcement. Extinction is technically defined by the withdrawal of reinforcement. In contrast,

To generate a high level of momentum for a specified class of behavior, arrange a high rate of reinforcement for a high rate of responding in the training situation. Once the behavior is established, it should persist effectively in the situation correlated with the rich schedule of reinforcement, even when reinforcers are withheld for a time. (Mace et al., 1992, p. 658).

Thus behavioral momentum may be similar to schedule effects, particularly given Mace et al’s., definition which indicates that “reinforcers are withheld for a time” implying that they are reinstated again. This would suggest a special, but not untrainable schedule. Behavioral momentum is typically used in special education as an instructional arrangement where a high rate of responding for one task is followed by the presentation of a low probability task. Experiments by Mace et al., (1988) indicated that behavioral momentum is a useful strategy where compliance for task completion “carries over” to over requests that are normally resisted. In any case (behavioral momentum as
a phenomenon versus schedule effects or behavioral momentum as an instructional arrangement) the essential empirical issue is that a response class can be made to persist at a rate for a time following the withdrawal of reinforcement.

Mace et al., (1992) have taken the notion of behavioral momentum and used it as an framework to explain behavior in sport settings. Mace et al., investigated 14 college basketball games during the 1989 NCAA season. The dependent variables observed were classed as : reinforcers (e.g., points, favorable turnovers), adversities (e.g., missed shots, unfavorable turnovers), and responses to adversities (favorable or unfavorable outcomes of the first possession of the ball following adversity). Data were analyzed using the number of events recorded as reinforcers during the 3-minute period preceding each adversity and dividing by 3 minutes to produce a rate per minute. Reinforcement rates were grouped into three classes. 0 or 0.3, 0.67 or 1.0, and ≥ 1.3 per minute. Figure 3 shows the percentage of adversities that a team responded to favorably as a function of the teams rate of reinforcement during the 3 minute preceding adversity. The data presented in Figure 3 indicate that when reinforcement was 0 or 0.3 prior to adversity the next possession of the ball was observed an average of 44% of the time. This value increased from 53% to 68% as the rate of reinforcement increased (from 0.67 or 1.0, to ≥ 1.3 per minute).
Figure 3. Percentage of adversities responded to favorably by the target team as a function of the rate of reinforcement (Mace et al., 1992).

Mace et al., draw the following conclusions from this correlational (but not experimental) data.

1. A team's favorable response to an adversity greatly increased as the rate of reinforcement increased 3 minutes prior to an adversity.
2. Basketball coaches called time-out from play when being outscored by their opponents an average of 2.63 for every 1 time-out called when the team was ahead.

3. Calling time-outs from play appeared to be an effective intervention in reducing an opponents rate of reinforcement. Rates of reinforcement during the 3 minutes following a time-out were nearly equal for both teams. (p. 657). Time-out may be viewed as an attempt to stop or slow down the pace or momentum of offensive plays. This is almost identical to Kounin's (1970) descriptions of movement management which he described over 2 decades ago. Movement management consists of two properties momentum (the pace of the lesson) and smoothness (the absence of breaks and stops in the lesson). Movement management can be influenced by two categories of "movement mistakes" slowdowns (teacher produced errors that slowed down the pace of the lesson) and jerkiness (breaks/stops in the pace of the lesson). Mace et al's., work in basketball, parallel's Kounin's (1970) work in the classroom. Though the goal of coaches and players in using time-out is very different than the teachers (winning versus instruction) who have a lot of slowdowns, the effect is the same (the pace of the activity, either game or lesson is slowed down).

This study represents the first study to address rate of reinforcement and rates of behavior in sports settings. It is an interesting beginning to a line of inquiry that has embryonic beginnings elsewhere as well (in special education and in the
nonhuman animal laboratories of behavior analysts). This methodology if applied to practice sessions or physical education classes (perhaps incorporating Kounin's definitions as well) might yield valuable insights into how the pace of activity is developed and hindered in physical education. The problem if any lies in operationalizing the definition of reinforcement. Precise specification and verification of the reinforcer (consequence) is a necessary requirement for the establishment of functional relations and while Mace et al., have presented an interesting new approach the study of games, it is a nonetheless a descriptive approach, and not yet experimental.

**Experimental Studies**

Bischoff (1982). In this study Bischoff investigated
1. If equal opportunities occurred in a team activity (volleyball) for males and females.
2. The success rate of students during game play.
3. How satisfied females and males were in participating on their various teams and who they would select as their future team mates.

Eighty-six students (40 females and 46 males) from three 7th grade physical education classes served as subjects. The students were observed for 30 games conducted over 10 days. For each game the female to male ratio was manipulated. There were three games played with a 5:1 ratio, three with a 4:2, seven with a 3:3, fifteen with a 2:4 and two with a 1:5 ratio.
The dependent variables included the total number of individual hits and serves, and the successes (outcomes) of each. Additionally, the proportion of individuals' hits and serves (females to males) was examined. Data were only collected during game play. The data were analyzed using t tests for independent groups to test for differences between the means of each group (males and females) and their total hits and serves, and successful/unsuccessful performances. A chi square test for independence was used to analyze the proportional data (male to female ratios).

The results indicated that males obtained significantly more hits than females, \( t(339) = -2.06, p<.05 \) and were more successful in those hits \( t(339) = -3.16, p<.01 \). There was no significant difference between groups on the number of serves, however males were more successful in their serves than females \( t(339) = -2.20, p<.05 \).

The proportional data analyzed by female to male groupings in games (e.g., 1:5) indicated that the grouping which produced the most equitable opportunities was the 3:3 (girls:boys) ratio (no significant difference found). Females also received a proportionally fair share of hits when the ratio was 4:2. Whenever the ratio placed more males on the team than females (i.e., 1:5 and 2:4) the females did not receive a proportional share of the hits.

Bischoff concluded by recommending 3:3 ratios in physical education be achieved by equal heterogeneous groupings. The
goal of achieving 3:3 (girl: boy) ratios in physical education classes is impractical, since class membership in terms of gender and numbers cannot be controlled. The study has however, several nice features. The selection of an independent variable which manipulated the players on a team (in this case by gender ratios) paved the way for similar manipulations by ability or skill level and by manipulating the size of the team (e.g., instead of regulation 6 on 6, what are the effects of 3 on 3, and 4 on 4, on variables such as opportunity to respond analyzed by gender and skill level). An additional and desirable feature of this study was the use of measures of success and outcome. The purpose of games is to score points. Evidence of the ability of teams to score points and the errors that prevent points being scored are valuable additions to (a) the feedback that a teacher can provide students and (b) the future instruction to be arranged by the teacher.

Brown (1986). [Note: This elementary study included here because of the paucity of games studies, and because it is representative of at least 3 other studies, Davis, 1990; Lawless, 1984; & Parker, 1984; conducted in similar manner in similar settings]. Brown examined the relationships between selected games (volleyball and soccer) modifications and student criterion process variables: ALT-PE and OTR's. The study sought to provide empirical evidence for the notion that modified games are more appropriate than regular dimensioned games. Brown
observed six students (per class) in two 5th-grade classes for 4
days per week over 7 weeks. The classes met for approximately
25 minutes per day. Systematic observation using coding
instruments designed to measure ALT-PE and OTR's were used to
capture the dependent variables. The modifications for
volleyball used a common set of rules and involved changing the
court and team sizes (a) one game, 1/2 the class on each team,
used the entire playing area; (b) two games, six on a team, each
court used 1/2 the playing area; (c) three games, five on a team,
each court used a 1/3 of the playing area; and (d) four games,
four on a team, each court used a 1/4 of the playing area. The
soccer modifications involved the same playing area and
common rules, but each modification assigned players to
different field zones. For example, when no zones were used
players could move anywhere. In contrast, when 1/4 zones were
used one player from each team played in each of the four
quarters.

Academic learning time-physical education data was
reported as a percentage of the total lesson time while OTR's
(including acceptable and unacceptable responses) were
reported as rates per minutes. Three males and three females in
each class were used as target students (n=6). Each of the
players was categorized as either high, moderate or low skilled.
A multielement design was used to analyze the data. The
modifications/treatments were systematically counter
balanced throughout the duration of the study.
The results indicated that no significant differences appeared between the modifications across sports, skill level, gender, using OTR's and ALT-PE. At least two important conclusions can be drawn from Brown's non-significant findings. The first is that either the differences between the modifications were not significant or that the measures used to determine the differences were insensitive or unable (i.e., other measures may account for the differences) to discriminate among modifications. Brown cited studies in his review of literature where modifications did and did not distinguish between student opportunity to participate. This issue remains unresolved in the contemporary literature as well (Brown's study was conducted in 1983-1984).

A second interesting product of Brown's investigation regards the similarity of the responses. Brown reported that in general data were remarkably similar and that differences were not apparent on the basis of ability level or gender of the students. This finding suggests that teachers' used a "one size fits all" approach to their instruction. That is the teachers failed to treat students differentially. If true the relevant question is, "Is the lack of differential treatment a product of the modification/s or the teacher instructional behavior?" At least one study (Ward, 1990) would support the conclusion the modifications were not responsible for the lack of differential treatment by the teacher, and that differential treatment by the teacher occurred during non-modified settings as well.
Comments

1. A problem faced by this study lay in defining a "normal game condition" for these classes. Some modification/s may have already been in place during the "normal condition." Brown attempted to deal with this by making a regulation game part of the treatment.

2. Brown reported that rates were calculated using total game time (from the moment of the first serve to the end of the class), not actual time the ball was in play. The overall effect would be to deflate the rate measures (an effect Brown acknowledged). Producing lower response rates than there actually were may have contributed to the non-significant findings. This issue is a difficult one. In educational settings the unstated goal much of the time is maximum activity. Yet in competitive environments this is not necessarily a goal. For example at the conclusion of a game where one team is slightly ahead in points of another the goal of the leading team will be to slow down the play. Alternatively if the leading team does not have the ball, but the opposing does and is "on a roll" then a useful defensive tactic is to call time out to break the pace of the offense. To use another example, in anaerobic activities (e.g., field events such as shot put, javelin and sprint events) more time is spent resting (interval training) than is spent in activity. When educational settings attempt to duplicate competitive environments "How is activity time to be calculated when wait time is a legitimate part of the strategy
or training regimen?" Using this logic (that wait time is a legitimate part of the game/activity) one must argue in favor of beginning to end measures. In this sense Brown’s rate measurements would be judged as legitimate!

Buck and Harrison (1990). Buck and Harrison conducted a study during two semester-long (22 days) undergraduate volleyball classes. Fifty-eight students arranged into two classes were pre-tested (A.A.H.P.E.R.D. Volleyball skill tests and the Stanley spike test) and grouped according to their performance (i.e., high-, medium-, low-skilled groups). The groups were determined relative to each skill. Thus a student might have found herself in the low skilled group for serving and in the high skilled group for setting. Using successful (a single judgment including topography and outcome), unsuccessful, and total trials as dependent variables they assessed the performance of students for the skills of setting, forearm passing, serving, and spiking. A one-way analysis of variance was used to determine differences between classes and between ability groups for contacts per serve (e.g., total sets/total serves) and the percentage of successful trials. All data were collected during game play and the authors reported that game play constituted the largest component of the instructional unit. With few exceptions the form of the games was a six-player arrangement on a regulation court.
There were no significant differences between class means, but there were significant differences between the means of the three ability groups for contact per serve and the percentage of successful trials. Buck and Harrison identified six relevant (to the present study) findings:

1. Most of the time in each class was spent in game play. This was not by design. Buck and Harrison reported that they did not provide any feedback or suggestions to the instructors regarding the instructional design of the classes.

2. For all ability groups the contacts per serve did not vary during the semester, nor did the percentage of successful trials. Buck and Harrison concluded that using game play to teach skill did not increase the students' performance level. The authors commented

   In fact, in viewing the videotapes, it was observed that as the emphasis turned to strategies and especially to running a 6-2 offensive, the level of play decreased. When these same students played after class, they employed no formal offense, the number of total contacts per serve appeared to increase, and there were more pass-set-spike sequences” (1990, p. 47)

3. “The low-skilled group contacted the ball fewer times than the other groups and was not as successful when contact was made.” (p. 48)

Buck and Harrison concluded by calling for more studies investigating a learning strategy that increases game-like play
to increase the number of successful trials for players (particularly the low skilled). Additionally they called for a learning strategy that improves the number of successful trials during non-game settings, presumably to develop skillfulness in players prior to game play, especially for the low skilled players.

Summary

This review of literature has examined the use of rate of response in educational settings. The chapter began with a discussion and definition of rate of response. A review of the empirical literature of education (previously summarized) revealed that rate has been frequently referred to in the literature, but not defined as events per minute. Responses have been qualitatively described as successful or incorrect (errors), and rate of response is conspicuous by its absence as a dependent measure, though it has received frequent mention in the anecdotal records and discussion of several of the studies.

In the physical education literature rate of response has been investigated most often as a student process measure. It has not been directly manipulated in any of the studies reviewed, thought the need to improve the amount of responding in physical education has been a recurring theme in many of the discussion sections/chapters of the studies reviewed.

In physical education rate of response is variable at the level of intra-class analysis when gender and ability are used to
differentiate responding and when rate of response is
differentiated by different content (e.g., volleyball versus
soccer). Rate of response was also variable when inter-class
analysis was used to compare differences between settings,
teachers and instructional tasks. There is evidence of
variability in the rate of response in physical education lessons.

The final section of this review examined the research
using rate of response in games settings. Investigations into
games research is still in its infancy yet several different lines
of inquiry have developed using process-product, mediating
process and behavior analytic paradigms. One way of viewing a
lesson is as a series of conditions or treatments imposed by the
teacher. In this view routines/review of past instruction, and
new skill development are seen as unique instructional events.
In this sense games are a special instructional condition and
event. The focus to date has been on skill development and
little work has been devoted to the development of strategic
play.

The task of this chapter in addition to describing the
literature was to find evidence that rate of response might be
an alterable variable. The nature and rates of responses do vary.
Neither the qualitative nor quantitative dimensions of the
response have been directly intervened upon in physical
education (though qualitative dimensions of a response have
been a principal focus in the general education literature in the
mastery learning models). There is however, considerable
descriptive-analytic as well as anecdotal evidence to suggest that qualitative and quantitative dimensions of responding in physical education can be intervened upon and improved.
CHAPTER III
METHODS AND PROCEDURES

This chapter is organized into four sections. Section one describes the methods of data collection involving machine observation and recording. In the second section, the collection of data using human observation and recording is discussed. The intervention is described in the third section. In the fourth section the research design and methods of data analysis used to answer the experimental questions are presented.

Arranging Machine Observation

The general strategy pursued in this investigation was to capture as complete a record of the events under investigation as was possible. Videotapes were used to produce a permanent record of the ongoing events that occurred in the gymnasium. The use of videotapes assumes that the events observed were accurately transduced to magnetic code and that this process was limited only by the "sensitivity of the lens" (e.g., the focus and zoom characteristics of the camera). Human observation was used to transduce these events into defined response classes. This section of the chapter addresses facets of the
investigation concerned with both the logistics of arranging for
machine observation at the site and the machine observation
itself.

Selection of Setting and Subjects

The School

The school selected for the investigation was a suburban
public high school with a population of 650 students from lower
socioeconomic income families. The school was chosen because
of a requirement of this study for a collegial relationship
between investigator and teacher/s. The investigator had
previously established a collegial relationship with the school in
his role as a university supervisor of student teachers. As such
the investigator was familiar with the school, its principal, its
schedule, its policies and its physical education teachers.

The Teachers

Two teachers participated in this study. One teacher was
selected from a pool of cooperating teachers used by The Ohio
State University School of Health, Physical Education and
Recreation. The criteria for selection included the teacher’s
reputation determined by recommendations from Professors, her
principal, and fellow teachers; 11 years teaching experience, and
informal observations of her teaching behavior over the previous
quarter by the investigator. The second teacher who served as a
subject in this study was a colleague of the first and was
employed half time as a physical education teacher and half time as a health education instructor. Her selection as a subject was based on the recommendation of the first teacher and because she was a professional peer in the high school selected for this investigation.

Both teachers were informed of the desire to conduct systematic observations of their classes and of the desire to intervene on the "instructional" conditions of their classes. The teachers were informed of the experimental conditions to be used, but not of the dependent variables to be observed. Their cooperation was sought in refining the experimental conditions for use in their classes. This consisted of several meetings to discuss the experimental conditions and two formal trials of the experimental conditions to identify any unanticipated problems with the experimental conditions. In return for their participation throughout the investigation, each teacher received three graduate credit hours.

The Setting

The investigation commenced two weeks into a new semester with two classes (grades 9 & 10) that attended physical education five days per week for 50 minutes per lesson. The 9th-grade class, served as the primary focus for this investigation. This class was observed for nine weeks. Data were collected for eight of the nine weeks (the first week was
devoted to a fitness unit). The instructional units consisted of a 13 day volleyball unit, a 15 day basketball unit, and a 10 day badminton unit. The second class, 10th-grade, served as a setting in which the intervention conducted on the 9th-grade class was replicated. This class was observed across the same instructional units occurring on the same day as the 9th-grade class.

**The Target Students**

Four students from each class were chosen for observation. The same students were observed for the duration of the investigation. Previous research has indicated that not all students respond the same in physical education lessons and thus the selection of target students is an important experimental consideration. Gender and skill level are variables commonly cited as responsible for differential responding (Buck, Harrsion, & Bryce, 1991; Dunbar & O'Sullivan, 1986; Siedentop & O'Sullivan, 1993). Son (1989) in his study of task congruence found that high, average and low skill students not only differed in their response rates, but in their performance of the stated task.

The selection of students who vary markedly in their response patterns in physical education was used by Lund (1990) as a strategy to exploit the variability between high and low achievers. From a teacher compiled list of high and low achievers in each class, four target students were selected (two
high achievers and two low achievers). Lund found that this strategy allowed her to observe and record the greatest variability in response patterns.

A similar strategy was employed in this investigation. The teachers were asked to select from their classes those students who were either high or low skilled students, and to identify frequent absentees. Two students from each group who regularly attended class were selected as target students without the knowledge of the teacher. Each pair of students included a male and female student to determine if gender was a variable for differential responding. In the event of a possible absence, at least one of the high or low skilled subjects were present for data collection.

Prior to the start of the study the teacher was asked to assign half of the class to one half of the gym. Students were directed and required (by the teacher) to remain in that half of the gymnasium for the duration of the instructional unit. Graham (1986) used a selection protocol to avoid selection bias by the teachers when students were assigned to each half of the gymnasium. The following protocol was derived from Graham's work and was used to guide decisions made by the teacher in selecting the class members to be at each end of the gymnasium.

1. High and low achievers must be distributed evenly between the two groups.

2. A general balance of males and females must be assigned to each half of the gymnasium.
When the teachers compiled the list of high and low skilled students in their classes and indicated which students would be at each end of the gymnasium, the target students were chosen.

**Gaining Entry**

After initial discussions with the teachers regarding their interest in participating in the investigation, permission to conduct the investigation was sought from the school principal. A cover letter (See Appendix A) and a prospectus was prepared and presented to the principal prior to a formal meeting to discuss the investigation (See Appendix B). The prospectus included a brief description of the investigation, as well as an outline of the methodology, and the proposed timetable. The investigator was already known to the principal, and to the teachers in his role as a university supervisor. Once permission was obtained from the principal, permission to conduct the study in the local school district was sought and obtained from the district superintendent of the local school district.

This research was exempt from review by The Ohio State University Human Subject Review Committee (exemption protocol number 92B0276Z). Exemption was granted by the Human Subject Review Committee because the investigation was conducted in an established educational setting and involved normal educational practices. In the case of this study, comparisons among instructional techniques constitute a legitimate investigation of existing gymnasium practices. The
identity of the teacher, the students and the school will not be revealed in any publications or documents, and no investigator was required to have any contact with the students.

Videotaping Procedures

All classes were conducted inside the same gymnasium for the duration of the study. Classes were videotaped using a video camera. The camera was located in the same corner position for each observation session. Each teacher wore a wireless clip-on microphone. The microphone was placed on each teacher prior to the start of their lesson. Taping began at the formal start of the lesson. If the scheduled time was 7.35 to 8.25 am the taping commenced when the class lined up to begin the lesson. Taping concluded when the teacher instructed the class to move to the change rooms.

Camera Organization

The following strategy for camera organization was used:

1. Aim to keep the image of the four target students, and the teacher and as much of the class as possible on the monitor.

2. When this was impossible select the image to be placed on the monitor from the following hierarchy: the four target students and the teacher, the four target students, or the same three target students for the duration of the lesson.
Subject Reactivity

Subject reactivity refers to the possibility that observed changes are not the result of the independent variable, but are due, instead to the presence of an observer and/or the measurement devices (Barlow & Hersen, 1984; Johnston & Pennypacker, 1980). The general strategy was to reduce the degree of obtrusiveness and to maintain a regular presence throughout the investigation (Barlow & Hersen, 1984). In this study the following tactics were employed to reduce the degree of obtrusiveness to the students and teachers:

1. In the week prior to the commencement of the study (during a fitness unit) the investigator with the video camera were present in the class as a “fixture” of the semester’s physical education lessons for the two classes being investigated. The presence of the investigator and the camera also allowed the video taping procedures to be piloted which helped reduce potentially distracting corrections once the data collection began.

2. Being present prior to the start of the lesson and leaving after the lesson thereby ensured that the researcher’s presence did not physically disrupt the lesson.

3. Conversations between researcher and students, and researcher and teacher (during the lesson) were not encouraged.

4. The camera and the investigator remained outside the designated lesson area and in as unobtrusive a position as
possible without interfering with camera len's view of the lesson.

5. The investigator's presence was explained to all students in each class. The students were informed (by the teacher) that the investigator was videotaping a number of physical education classes at the school over the semester, for the general purpose of observing the work of teachers and students.

6. The teachers were asked if they knew of the identity of any of the target students midway through, and at the end of the investigation. Neither teacher was aware of the identity of the target students midway through the investigation. When asked at the end of the investigation Ms. East (Teacher: Class 1) correctly identified one student, but Ms. West (Teacher: Class 2) remained unaware of the identity of her target students.

The presence of others (e.g., student teachers and University supervisors) during physical education lessons was not an uncommon occurrence in this school. In this sense it was a common occurrence to have others present during physical education lessons.

Arranging Human Observation and Recording

In this investigation a camera observed and a machine recorded (video tape) the live events. Once the videotapes were obtained they were observed, and using systematic observation instruments the data were recorded.
Selection of Response Classes

Behavior does not occur in a vacuum: it is preceded by its history and surrounded by other ongoing events. Behavior in applied settings cannot be adequately studied without reference to its context. Kantor's (1959) call to look for context beyond the three term contingency behooves behavior analysts to examine the environment not just in terms of a three term contingency but in terms of the surrounding ongoing events. The terms antecedent and context are synonymous. They both refer to the occasion upon which an operant (behavior and consequence) is emitted. However a caveat is warranted. Just because events are correlated with each other (e.g., a particular prior occurring event and a response), does not make them functional. Conversely events that have happened in past lessons (e.g., teacher knowledge of learner performance) are not generally accessible to the observer of the current conditions, but this does not make them less functional.

An explanation of an operant occurring in its context is limited in the absence of a functional relation (y occurs, if x occurs), to a correlation (y and x occur at the same time). The specification of a response class often begins with repeated observations of the setting under investigation. Recurring behavioral events are observed, these events are categorized as conspicuous behavioral events occurring in that particular setting. This inductively derived list then forms the basis of a formal description of each behavioral event. An alternative method is to
extrapolate the definitions used in one setting to another, and through repeated use, modify the definitions to include the events of the new setting. This method is often used to modify instruments and behavioral categories. Both methods yield descriptions of the response classes in terms of their antecedent and consequent events (descriptive and correlational, not functional).

Several researchers in physical education over the past decade have used a contingency (antecedents, responses, and consequences) framework to explain the instructional events that occur in the gymnasium where the task was the unit of analysis. Alexander (1982) and Tousignant (1982) first used a contingency framework to explain tasks and the performance of those tasks in a physical education lesson. Since then, successive investigations (Griffin, 1991; Lund, 1990; Marks 1988; Son, 1989) have refined the framework. Figure 4 presents the current relations as reported by Griffin (1991) and Son (1989) using a diagram of the contingency and indicating an analysis occurring at two levels. The first level involves the congruency between the stated task and the actual student response. The second level requires some judgment (correlation) about the controlling consequences of the response. Several investigations (Griffin, 1991; Jones, 1989; Lund, 1990; Marks, 1988; Son, 1989) have used descriptive/correlational techniques to report on the contingency relations in the task system. To date there have been no functional analyses conducted in physical education.
In this investigation the contingencies were experimentally analyzed in terms of antecedents, responses, and consequences. Figure 5 diagrams the categories used in the analysis of the instructional contingencies in this study.

**Antecedents**

The antecedent events in this study are examined relative to the lesson component, the type of task, and the task specification.
<table>
<thead>
<tr>
<th>Context and Antecedents</th>
<th>Task type</th>
<th>Task specification</th>
<th>Student response</th>
<th>Response product</th>
<th>Teacher behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warm up</td>
<td>Instructional</td>
<td>Implicit</td>
<td>Correct</td>
<td>Successful</td>
<td>Redirect</td>
</tr>
<tr>
<td>Skill practice</td>
<td>Cognitive</td>
<td>Partially explicit</td>
<td>Incorrect</td>
<td>Unsuccessful</td>
<td>Support</td>
</tr>
<tr>
<td>Scrimmage/Routines</td>
<td>Informing</td>
<td>Fully explicit</td>
<td>Modified upwards</td>
<td>Completed</td>
<td>Correct</td>
</tr>
<tr>
<td>Game</td>
<td>Extending</td>
<td></td>
<td>Modified downwards</td>
<td></td>
<td>Intensify</td>
</tr>
<tr>
<td>Management</td>
<td>Refining</td>
<td></td>
<td>Other</td>
<td></td>
<td>Present</td>
</tr>
<tr>
<td>Transition</td>
<td>Applying</td>
<td></td>
<td>Unfair opportunities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>Routine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. The categories used in the analysis of the instructional contingencies
Lesson components.

Using descriptive categories initially developed by Siedentop, Tousignant, and Parker (1982) the lesson component requires the observer to determine "What is the class as a whole doing"? In this category there are six descriptors. These are:

Warm up: Time devoted to exercises designed to ready the class participants for the lesson content.

Skill practice: Time devoted to activities whose purpose is to develop skill.

Scrimmage/ Routines: Time devoted to generalization of skills to an applied setting, but where there is frequent instruction and correction occurring e.g., teacher feedback during a 5 on 5 basketball game.

Game: Time devoted to situations where the class is practicing in a competitive situation without intervention from the teacher for the purposes of instruction e.g., a volleyball game.

Management: Time devoted to class business that is unrelated to instructional activity e.g., taking roll.

Transition: Time devoted to managerial and organizational activities related to instruction e.g., changing stations.

Knowledge: Time devoted to describing and presenting information about lesson content i.e., instruction

These data were summarized into three broad classes: general content (including warm up, management and transition), knowledge, and activity (including skill practice,
scrimmage/routines and game). The data were recorded as duration measures and then converted to a percentage of the total time of the lesson.

**Task type.**

Rink (1985) categorized tasks according to their purpose in progressing the skill toward the terminal goal. For example tasks that introduce the student to the content are called "informing" tasks because they inform the student about a new task. This study further modified Rinks' categories by drawing a distinction between teacher instruction and cognitive tasks to better reflect the difference between listening to instructions and participating in a cognitive task.

**Instruction:** Any explanation or demonstration of a skill or strategy or the background thereof where the student role is to attend to the presentation. Instructional tasks require only that the student be judged as attending or off-task. Instructions are not stated as tasks, but they are often linked to activity tasks that follow them. To be coded as an instruction, the orientation of the class has to be clearly an instructional mode, rather than a brief/demonstration that might accompany a refining or extending task.

**Cognitive:** Tasks that require students to "think" and/or "problem solve." For example a cognitive task might be stated as follows "gather in your four person team and derive an offensive strategy for basketball, complete a simple diagrammatic
description of the strategy on these sheets and turn them in to me—you have 10 minutes—go."

Informing: Initial presentation of information about how to do a particular task (e.g., the first task related to forearm passing in a volleyball unit).

Refining: Tasks that focus on qualitative improvement of the skill/strategy. Conditions for practicing do not change. Only the focus on the task practice changes (e.g., continuing to practice the forearm pass as described in an informing task, but with special focus on bending the knees).

Extending: Tasks that change the conditions of practice to alter the focus of skill development, typically to make it more complex or difficult (e.g., changing the conditions of the forearm practice task so that it is more difficult, such as in moving partners further apart or placing them on either side of a net).

Applying: Tasks that require practice in situations similar to those in which they were used in games or other performance settings. Applying tasks typically have a clear goal and criterion to judge success or completion.

Routine: A task that has been described and practiced in previous lessons and is instituted typically by a shorter task cue (e.g., "do your free throws" as a cue for students shooting 5 free throws each, during each lesson of a basketball unit).
Task specification.

Mager (1975) defined task statements (which he called objectives) as requiring three characteristics to be considered explicit. Explicit tasks a) describe the situation that is the condition/s under which the task occurred (e.g., serve “from the baseline”); b) state the task or behavior to be performed (e.g., serve, kick, listen, write); and c) provide a criterion against which the performer (or teacher) can evaluate performance (e.g., accuracy, speed, number of successful trials).

Task statements have been analyzed in terms of their degree of explicitness (Griffin, 1991; Jones, 1989; Lund, 1990; Marks, 1988; Son, 1989). The degree of explicitness is considered a function of the presence of the characteristics of task statements. The following categories were used to describe the explicitness of the task statements made by the teachers in this study:

Fully explicit task statements: A task description that includes information about the conditions under which to practice the task, the task itself, and some criterion by which success and/or completion can be judged. A task statement is judged fully explicit if it includes all three of these elements (the task, conditions, and criterion).

Partially explicit task statements: A task description that excludes information about one of the components of a task statement. For example stating the task and the conditions for practice, but not giving a criterion for success or completion.
Thus a task statement is judged partially explicit if it includes 2 of the 3 elements.

Implicit task statements: A task description that lacks specificity, but implies a set of conditions or a criterion (e.g., requiring students to shoot baskets, but not specifying either how many baskets to shoot or what technique they should use). A task statement is judged Implicit if it includes only 1 of the 3 elements.

The actual statement made by the teacher was captured on the video tape through the use of a wireless microphone. The observers of the videotape recorded an abbreviated version of the task statement on the coding sheet in the notes column.

Response/s

Student responses.

Student responses were categorized according to their topography judged by the observers to be either correct or incorrect, modified upwards or downwards, or responses that are not defined by the previous descriptors e.g., off-task behaviors. Barrett, Johnston and Pennypacker (1986) noted that the "definition of response classes for assessment should be sufficiently molecular to permit measurement of the variability necessary for sensitive analysis and successful treatment" (p. 167). A problem arises in the specification of response classes for skill topography. If the teacher does not, or fails to specify the response class in a sufficient manner to allow for clear
discrimination by the investigator, what criterion should be used to judge the correctness of the topography? The only justifiable position requires an assessment by the observer of the particular performance of a student relative to an "acceptable working form" that is of sufficient quality to meet performance requirements of other teachers (in this case the observers). The response classes for topography are:

Correct: The response is topographically correct as judged according to the criterion stated in the task, or was an appropriate response that indicated an "acceptable working form."

Incorrect: The response is topographically incorrect as judged according to the criterion stated in the task, or was an inappropriate response when judged as an "acceptable working form."

Modified upwards: The response was more difficult than the task demanded.

Modified downwards: The response was less difficult than the task demanded.

Other: Responses that do not meet the other category descriptions such as off-task behavior.

Unfair Opportunities: An event that was designed to produce a response, but which placed the student at a disadvantage through no fault of his or her own (for example a misfed ball).
These variables are reported as a rates of response relative to a) the phase of the lesson (e.g., skill practice versus game) and b) the nature of the antecedent-operant relation (i.e., free operant versus restricted operant).

Consequences

Behavior is commonly defined in terms of a three term contingency. Operants however are defined as a relation between a response and a consequence. The antecedent conditions may set the occasion for responding, but consequences maintain or change the responding. Consequences are defined only by their effects and may be discriminated from subsequent events in this way. This discrimination allows for analysis of the function of the subsequent events. This study has described the subsequent and functional events (consequences) following the student response (topography) in terms of response products (the change in the environment as a result of the student performance i.e., the effects of the throw) and in terms of teacher behavior.

Response products.

Successful: The outcome of the response is judged successful according to the criterion stated in the task or was of sufficient quality to have met the criterion of the activity in a sport situation (e.g., the ball went over the net, landed in).

Unsuccessful: The outcome of the response is judged unsuccessful according to the criterion stated in the task or was
not of sufficient quality to have met the criterion of the activity in a sport situation (e.g., the ball hit the net, landed out).

Completed: The response product could not be described using the above categories, but was completed (e.g., dribbling around a set on cones without losing control of the ball).

Incomplete: The response product could not be described using the above categories and was incomplete (e.g., not completing the specified task or losing control of a ball while dribbling around cones).

Teacher behavior.

Teacher behavior is a class of behaviors whose central purpose is to arrange conditions for instruction. However teacher behavior can function as an antecedent or as a consequence. Teacher behavior could serve both functions if it were considered part of a behavior chain. For example a task statement made by the teacher might function as reinforcement for completing the previous task correctly and concurrently set the occasion for a new or modified response. Thus a caveat is necessary. Though teacher behavior is categorized as a consequence in this study for the purposes of coding, its potential function as an antecedent stimulus event should not be ignored. Five response classes have been used to describe the behavior of a teacher.

Redirect: Teacher behaviors where the purpose is to prohibit certain behaviors (usually misbehavior) in favor of
others (e.g., on task behavior). A desist would be a typical example of redirection.

Support: Teacher behaviors where the purpose is to maintain or improve the current conditions.

Correct: Teacher behaviors where the purpose is to modify or change student skill responses (e.g., shape a skill, or extinguish a response).

Intensify: Teacher behaviors where the purpose is to increase the rate of responding (e.g., prompts, hustles, high rate repetitive verbal instructions such as “go, go, go”).

Present: Occasions where the teacher is present (i.e., standing next to the student or the group that the student is part of), but does not emit any vocal behaviors or gestures.

Observation and Recording Procedures

Figure 6 illustrates the coding sheet used to record the variables under investigation in this study. The observation instrument was designed to code the variables from a videotape record where a running digital time readout was imprinted on the videotape tape. The instrument uses continuous recording of behavior to record three characteristics of behavior: the temporal locus (i.e., when in time), the temporal extent (for how long), and repeatability (how many times). These properties of behavior allow the data to be reported in terms of number, duration, frequency (rate), and celeration. The coding sheet was
divided into 10 columns (see Figure 6). The observational protocol is described as follows:

Column 1: Using the real time of the video-tape a record is kept of the start and finish of each episode. An episode is defined as a class event preceded by a task statement (including times when students are following previously established rules and routines).

Column 2: This column requires the observer to choose from the sub-categories describing the lesson component (e.g., warm up, skill practice) according to the definitions of each response class. If management or transition are coded no further records need be made on the coding sheet save for notes (see column 10).

Column 3: Similarly this column requires the observer to choose from the sub-categories that describe the type of task (e.g., routine, cognitive task) according to the definitions of each response class. Cognitive tasks require only that the student be judged as on the stated task (appearing to attend) or off-task. No further records need be made on the coding sheet save for notes regarding the context.

Column 4: This column requires the observer to choose from the sub-categories that describe the task specification (e.g., explicit, partial) according to the definitions of each response class. As well entries in this column require an abbreviated version of the actual stated task to be recorded in the notes column (column 10).
### Figure 6. An example of a completed coding sheet.

<table>
<thead>
<tr>
<th>Time</th>
<th>Lesson component</th>
<th>Task Type</th>
<th>Task specification</th>
<th>T/Student 1</th>
<th>Response/ product</th>
<th>T/Student 2</th>
<th>Response/ product</th>
<th>T/Student 3</th>
<th>Response/ product</th>
<th>T/Student 4</th>
<th>Teacher behavior</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08.00</td>
<td>WU</td>
<td>I</td>
<td>PE</td>
<td>C/U</td>
<td>C/U</td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>S</td>
<td>S</td>
<td>Volleyball</td>
<td></td>
</tr>
<tr>
<td>10.15.28</td>
<td>SK</td>
<td></td>
<td></td>
<td>C/U</td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>S</td>
<td>I</td>
<td>&quot;Groups of 2, set ball back and forth, fingers point at ceiling after set&quot;</td>
<td></td>
</tr>
<tr>
<td>10.17.35</td>
<td></td>
<td></td>
<td></td>
<td>C/U</td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>I</td>
<td></td>
<td>teacher monitoring individual FB to students.</td>
<td></td>
</tr>
<tr>
<td>10.19.39</td>
<td></td>
<td></td>
<td></td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>C/U</td>
<td>I</td>
<td></td>
<td>&quot;feed ball, set, set&quot;</td>
<td></td>
</tr>
<tr>
<td>10.24.28</td>
<td></td>
<td></td>
<td></td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>C/S</td>
<td>0</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Columns 5-8: These columns refer to the same target students in each class. Decisions about the correct and incorrect topography, and the response product.

Column 9: This column requires the observer to choose from the sub-categories that describe teacher behavior (e.g., redirect) according to the definitions of each response class.

Column 10: This column is used to record field notes, and abbreviated task statements relative to the lesson.

**Equipment and Materials**

To collect data (off line) from the videotapes, a video recorder, coding sheets, pencils and a freeze frame remote control were used. Tapes were replayed on a video cassette recorder that featured both fast and slow motion replay, together with clear freeze frame capabilities.

**Description and Training of Observers**

Two individuals collected and coded data for this study. One was the investigator. The other observer was a graduate student who participated in both data collection (videotaping) and interobserver agreement. All observers followed the training guidelines established by Siedentop, et al., (1982). These guidelines involve:

1. Learning the behavioral definitions of each category of the instrument.
2. Then checking these definitions against a training questionnaire where observers were required to match a descriptor with its definition.

3. When a criterion of 100% correct identification occurred observers were then asked to match the descriptor with a written behavioral vignette. A criterion of 90% correct identification was set.

4. When each observer met the criteria for acceptance in 3 videotaped vignettes were used in the same way as the written vignettes were used in #3. Again a criterion of 90% was set.

5. Finally each observer coded a continuous videotape to practice both coding and manipulation of the freeze frame control of the video recorder.

Following training, observers commenced coding of the videotapes collected in the study.

**Interobserver Agreement**

**Dependent Variables**

As scientists we employ several criteria to test the believability of data. The extent to which data meet these criteria, the stronger our confidence in that data. A common tactic is to use interobserver agreement (IOA) measures to answer questions concerning the reliability of observations (Harris & Lahey, 1978). The emphasis when IOA's are used is on finding an agreement between observers. The actual occurrence or non-occurrence of the behavior is not reported, only the
percentage of agreement. Increasing the number of observers does not in any way increase the confidence in the believability of the data. Measures of agreement no more assure the reader of accuracy than "if someone were to buy several copies of the morning paper to assure himself that what it said was true" (Wittgenstein, 1953, p. 53).

Interobserver agreement is necessary in the absence of a permanent record of the events under consideration. However when a permanent record is available accuracy should be used to gauge the confidence of the data. As Dugas (1983) succinctly comments:

When observers are coding live or in the field, some level of reliability or agreement between independent observers in training and field settings seems justifiable in deciding whether the response class definitions are appropriate. However, when coding is from videotapes, which represent a permanent product of the true value of the phenomenon, the focus should be accuracy of the data and not merely reliability of independent observers. (p. 56)

The goal of achieving a believable demonstration of events, rests with the ability of the researcher to collect accurate data, which represent "true" states of nature (Johnston & Pennypacker, 1980; Sidman, 1960). The collection of accurate data relies on both instrument and human observers being reliably calibrated across time. Human calibration is best conceived of as the repeated agreement of what data represent
(across time) by a scientific community. Accuracy requires a permanent product of the event to be kept (i.e., a video-tape) so that it can be referred to across time. The ability of data (coded from the video-tape) to stand the test of accuracy across time by many observers reflects Sidman's (1960) notion of the independence of data from the purposes for which the investigation was conducted. "Good data are always separable, with respect to their scientific importance, from the purposes for which they were obtained" (p. 3).

However, accuracy is dependent upon the observer's ability to discriminate and compare the observed behavior with response class exemplars. When the exemplars are not specified judgments are required. In this investigation the investigator did not intervene on the specificity of the task statement and as such could not guarantee that precise and explicit task statements would be delivered by the teacher, thus judgments about the response class (i.e., topography) serving as the dependent variables were necessary. Such judgments are reported as a percent of agreement between two observers using the following formula (equation 1).

Equation 1

\[
\text{Percent Agreement} = \frac{\text{Agreements}}{\text{Agreements} + \text{Disagreements}} \times 100
\]
Observer agreement checks were conducted on 30% of all data collected and distributed equally among baseline and each treatment. Interobserver agreement was obtained on 6 classes of dependent measures: Agreement and disagreement between observers regarding the duration and type of lesson component, task type (e.g., instructional versus cognitive), task specification (e.g., partially explicit or implicit), the number and type of student responses and the number and type of teacher responses is reported in Chapter 4. A criterion level of agreement between observers was set at 85%, below which the reliability of the data would be jeopardized.

Independent Variables

In most applied research the dependent variable is the focus of interest. Demonstrated experimental control of the independent variable and measurement of the dependent variable are necessary and sufficient conditions in order to claim a functional relation. Frequently, the independent variable is not observed and cannot confirm or refute its own variability. If the goal of science to explain variability (Johnston & Pennypacker, 1980) then this protocol does not constitute good scientific method. As Peterson, Homer and Wonderlich (1982) observe:

A curious double standard has developed in operant technology whereby certain variables (e.g., social behavior, smiling, and attention) routinely have operational
definitions and some measure of observer reliability when
the observed behavior is the target response or dependent
variable . . . but no such rigor is applied to the same
behaviors when they appear as antecedents or
consequences to the target behavior, as independent
variables . . . Either such precautions as definition and
reliability of observation are necessary or they are not.
The observational literature clearly suggests they are (pp.
478-479).

Procedural reliability is defined by Cooper et al., as “a
measure of the extent to which the actual application of the
independent variable over the course of an experiment matches
its description in the method section of a research report” (p.
239). This study reports the procedural reliability of the
independent variable in terms of each teacher’s implementation.
The fidelity of treatment implementation by the teacher was
judged using a task analysis of each treatment. The task
analysis consisted of a checklist of essential components of the
treatment (see Appendix C). Checks on the fidelity of the three
treatments were conducted on 30% of all data collected and
distributed equally among each treatment.
The Intervention

The intervention consisted of the addition of three instructional treatments to the existing class conditions found in the two physical education classes. The treatments were:

Treatment A, An untreated baseline: The existing instructional conditions of the class constitute the current treatment. This condition served as a baseline against which to compare the effects of the other treatments.

Treatment B, Hustle: Refers to an treatment where high rates of teacher statements were designed to increase rates of student responding. In their study of basketball's John Wooden, Tharp and Galimore (1976) reported on the importance of hustles and prompts that were used to intensify practice. They commented "His hustles as we call them, have a special value in physical learning . . . If such concentrated effort ever finds its way into the ordinary classroom, the results might be rewarding" (p. 78). This treatment was designed to substantially increase the number of teacher hustles. The criterion was determined relative to baseline measures. Treatment B required at least a fivefold (x5) increase over baseline in the number teacher statements designed to increase student responding to previously stated tasks. These verbal statements have in the past been categorized as prompts and hustles, or as repetitive cues such as "pass, pass, pass, keep it going, pass, keep going." Teachers were asked to maintain the hustle for at least 3 minutes, but generally not to exceed 5-6 minutes.
Treatment C, Sprints: Derived from Precision Teaching (Lindsley, 1991, 1992) and based on a strategy described by Hook (in press) this treatment involves the use of timed practice trials (sprints) where correct performances were counted, and where the focus on the instructional treatments is to produce as high a rate of responding as is possible for a particular context. Usually sprints were repeated and conducted on similar tasks during the lesson. Students were grouped in pairs and given one minute to perform as many repetitions as possible. Hook (in press) termed this repetition of sprints a "skill hustle." Self reporting by the students (verbal only) was used by the teacher to keep track of student performance. The teacher however selected a student or students at random and observed the performance/s (successes and errors) in order to hold students accountable for the accuracy of their self reporting. Students were praised and recognized for their high rates of successful responding. In this treatment the focus of the teacher's attention was on the performing student.

Treatment D, Peer sprints: This was an identical treatment to treatment C except that it involved a peer keeping a record of the performance of another student. This treatment involved students being grouped in pairs. Each couple had a cardboard sheet (on which the name of each student in the pair is written). Figure 7 is an illustration of a completed peer recording sheet. One member of the couple was required to perform a skill task for one minute in duration while the other kept a record of the
number of successful outcomes (i.e., balls through the hoop). At the end of the minute the other student then changed position with his or her partner and the one minute sprint began again. The use of peer recording is included because a previous study of accountability in physical education (Lund, 1990) identified the effects of formal accountability on skill performance, and because studies by Greenwood et al., (1988, 1989, 1992) have indicated that peers can be taught to effectively discriminate another's behavior, and that mere observation serves as a powerful treatment onto itself. In this treatment the focus of the teacher's behavior was on the recording student.

<table>
<thead>
<tr>
<th>Name: Andrea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials:</td>
</tr>
<tr>
<td>Mon</td>
</tr>
<tr>
<td>Tue</td>
</tr>
<tr>
<td>Wed</td>
</tr>
</tbody>
</table>

Figure 7. An example of a peer recording sheet.
The Tasks

The investigation consisted of three studies conducted in basketball, volleyball, and badminton respectively. In each study (setting) interventions occurred to boost the rate of response for the target students and data were recorded to report on the effects of this increase in responding upon the other dependent measures (tasks, time use). The three treatments used were hustles, sprints and peer sprints. This section of the chapter briefly describes the tasks used by the teachers and which were used to test the efficacy of the treatments.

Basketball During basketball data were collect on the effects of the sprint and peer sprints compared to baseline during dribbling tasks that required the students grouped in pairs, taking alternate turns, to start at the baseline and dribble to the mid point of the basketball court and return. The tasks involved students (a) dribbling to the mid-line and returning, (b) dribbling to the mid-line using crossover of the ball from hand to hand during the circuit, and (c) dribbling to designated places, stopping for 2 bounces, and then moving on. Due to the similarity of these tasks the data were reported as the mean for all three sub-tasks.

The hustles treatment was applied to two different tasks. The first task required all students to navigate around a circuit (consisting of cones and staying on basketball court lines) dribbling, and changing hands used for dribbling. During baseline for this task the teacher began with a arbitrary rate of prompts
and hustles, then during the intervention increased the rate of prompts and hustles. The second task used for the hustle treatment was a passing drill. In this case the students were arranged in lines of 3 with the first person in each line passing to the first person in opposite line. Students then quickly changed sides. Like the dribbling drill, the teacher began with a arbitrary rate of prompts and hustles, then during the intervention increased the rate of prompts and hustles.

**Volleyball.**

During volleyball only the sprint and the peer sprint drill were used. Students were again paired together and three different tasks were used. The first task was called the “wall drill” and consisted on the students bumping a volleyball continuously of a wall. The second task called the “fed drill” required one of the students to toss (lob) the volleyball so that his/her partner could return the ball using a bump or a set. The third drill “1V1” was a continuous bump and set drill where the goal was to keep the ball in play for as long as possible.

**Badminton.**

In this study both the “fed drill” and the “1V1 drill” were used. During the fed drill one student tossed the shuttlecock to the other student and a underhand clear or overhand clear was used to return the shuttlecock. During the “1V1” was a continuous play using under- and overhand clears.
By What Methods Might Answers to the Experimental Questions Best be Obtained?

The experimental questions in this investigation are comparative in nature. They compare differential effectiveness of several treatments on the dependent variables. The design choices in answering comparative questions are between-group comparisons and within-subject comparisons. Between-groups comparisons (e.g., ANOVA) are suitable for questions about populations where inter-subject variability is the focus of the investigation. However, the experimental questions in this investigation are questions about individuals i.e., intra-subject variability, and questions of intra-subject variability are the domain of within-subject designs.

Two comparative within-subject designs were considered for selection, the multi-treatment design (Tawney & Gast, 1984) and the alternating treatments design (Barlow & Hayes, 1979). Both designs face the problem of multiple treatment interference, the influence one treatment may have on the effectiveness or ineffectiveness of other treatments. Ulman and Sulzer-Azaroff (1975) describe multiple treatment interference as problems concerned with sequential confounding and those concerned with carryover effects. Sequential confounding occurs when one cannot distinguish between treatment effects because one treatment always follows another (e.g., A-B-C-A-B-C) and
the best one can conclude is that the treatment effects occur only after the sequence of preceding treatments. The standard strategy used to address sequential confounding is to counterbalance or randomize the order of treatments (Barlow & Hersen. 1984). This requires experimental periods of lengthy duration.

Carryover effects provide an example of changes in behavior under one set of conditions, which are caused by changes in behavior under a different set of conditions. When data paths after each treatment have a tendency to interact to produce changes in direction, the opposite of what was predicted the phenomenon is labeled contrast effects. In a similar manner when data paths after each treatment have a tendency to interact to produce a continuation of the effects of the previous treatments the effect is known as induction or generalization (Barlow & Hersen. 1984; Ulman & Sulzer-Azaroff, 1975). The multi-treatment design was rejected as a usable design in this study because of the length of time required to counterbalance treatments to address the problems of multiple treatment interference.

The alternating treatments design (Barlow & Hayes, 1979) is defined by the rapid alternation of treatments each associated with its own discriminative stimulus (i.e., treatment). In the alternating treatments design each successive data point (1) serves as a basis for the prediction of future levels of behavior under that treatment, (2) it provides verification
of previous predictions of performance under that
treatment, (3) it is a replication of the differential effect
produced by the other treatment(s) that are part of the
design (Cooper et al., 1987, p. 182).

The alternating treatments design was the design of choice
for this study because

1. It minimizes sequence effects through the rapid
alternation (random or non repeated sequences) of treatments
across session and within session (Cooper et al., 1987; Sidman,
1960). In so doing it reduced the need for lengthy periods of
experimentation.

2. The rapid alternations allows for rapid discriminations
between treatments which also serves to identify carryover
effects if they occur (Barlow and Hersen, 1984; Sidman, 1960).

3. There is no need to reverse or withdraw treatments to
demonstrate experimental control since the control is
demonstrated by the stimulus control of the various treatments
(Cooper et al., 1987; Tawney & Gast, 1984).

4. The variations of the alternating treatments design
include those which do not require a stable baseline or a baseline
at all to begin treatments, and those which allow for a “best
treatment” phase in the final stage of the design arrangements
(Cooper et al., 1987; Tawney & Gast, 1984).

Despite its utility several authors note caveats regarding
use of the alternating treatments design. These caveats can be
categorized regarding concerns about the effects of rapid
alternation on internal and external validity, and concerns regarding procedural integrity across sessions. Cooper et al., (1987) warn that while the rapid alternation of treatments affords many advantages it is not a "natural" condition. This requires that either the treatments have a previous experimental history in order to be confident of their long term effects and comparisons or that current study be replicated in order to establish generality. Johnston and Pennypacker (1980) caution that when across session treatments are used and where the total number of applications of each treatment are few, instability may be hidden and thus pose a threat to internal validity. Cooper et al., (1987) warn that despite the rapid alternations multiple treatment interference must always be suspect in an alternating treatments design. Both they (Cooper et al.) and Johnston and Pennypacker (1980) advocate a period of time where the condition is administered alone until stable responding is demonstrated.

The Design Arrangement for this Investigation.

This study used a within-session and a across session alternating treatments design which was replicated across three settings (volleyball, basketball and badminton) concluding the replication in each setting with either a best treatment phase or class choice phase. The treatments were also systematically replicated across a second class with a different teacher, but
using the same settings (the same school, the same gymnasium, and the same content areas).

The conditions are designated as:

Condition A: Baseline conditions
Condition B: Hustles
Condition C: Sprints
Condition D: Peer Sprints

(For specific details of the content of these conditions please refer to the details under the heading "The Intervention" in this chapter).

Table 5 indicates treatment sequence for the within-across session alternating treatments design, for the basketball study. The treatment order was not random, but repetitions of treatment sequences have been avoided. The design arrangement grants class choice of a favored treatment determined by a majority vote (See day 11 of Table 5).

Figure 8 provides an illustration of the graphic display of the above design arrangement using a within and across session alternating treatments design (for display purposes the levels of the dependent variables under each treatment were held constant, and only a single task comparison is displayed).
Figure 8. A model of the proposed alternating treatment design.
Table 5

Design Arrangement across Sessions for the First Setting

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A A B1 D B1 D A C C D C</td>
<td></td>
</tr>
<tr>
<td>B1 A A D C B2 A A A A</td>
<td></td>
</tr>
<tr>
<td>C B1 A A A2 D B2</td>
<td></td>
</tr>
<tr>
<td>D A2</td>
<td></td>
</tr>
</tbody>
</table>

Notes: A = Dribbling baseline; A2 = Passing baseline; B1 = Dribbling hustle; B2 = Dribbling hustle; C = Sprint D = Peer sprint. A served as baseline for all treatments involving dribbling. Though the tasks were different for hustles and the sprint/peer sprint treatments the response rates were almost identical during baseline.

Data Analysis for the Experimental Questions

This section presents examples of data and the graphical display tactics used to answer each experimental question

Experimental Questions

1. What are the effects of modified classroom instructional interventions in physical education?
1.1 Do the treatments improve the response rate, success rate, and correct topography rate of the target students relative to “standard” practice?

1.2 Are any of the treatments superior to the others relative to response rate, success rate, and correct topography rate?

1.3 In a forced choice between treatments which treatment is selected by the students?

1.4 Which (if any) of the treatments are used by the teachers in their other classes?

Questions 1.1 and 1.2 were answered by a visual inspection of the alternating treatments design (See Figure 8). Question 1.3 was determined on the 11 day of the study when students were asked what treatment they would prefer and question 1.4 as reported earlier (see social validity) was determined through the use of anecdotal records throughout the study provided by informal observation of other classes taught by the teachers.

2. How does an increased rate of responding affect the behavior of students and teachers in physical education lessons?

2.11 Does an increased rate of responding affect high and low skilled students differently?

2.12 Does an increased rate of responding affect male and female students differently?

The above questions were answered by a visual inspection of the alternating treatments design.
2.2 Does the explicitness of tasks stated by the teacher change as the rate of responding changes?

These data were analyzed by looking for changing patterns of explicitness during the treatments when compared to baseline.

2.3 In what ways does an increased rate of responding affect how students and teachers spend their time in physical education lessons?

Using a graphic display of time categories (general content, knowledge and activity), a within and across lesson comparison was made by visually inspecting the data.

Social Validity

One of the defining characteristics of behavior analysis is its emphasis on socially important behaviors (Baer, Wolf & Risley, 1968, 1987). The criterion used has been that the behavior was important to the subject or to the society (Baer et al., 1968, 1987; Cooper et al., 1987). The approach has been a pragmatic one. The treatment/s had to work, and the reduction, increase or maintenance of behavior needed to be measured not in terms of how much change occurred, but rather how much was needed (to be member of society).

Wolf (1978) provided an initial strategy for use in evaluating goals (does this behavior need to be changed), in evaluating procedures (are these procedures used in society) and in evaluating effects (are these behaviors demonstrated by other
members of society) of applied behavioral research. This strategy used tactics that were subjective in nature involving the use of questionnaires and interviews as well as opinions from experts or significant others who interacted with or observed clients during the treatment/experiment. The implicit assumption made by Wolf (1978) was that self report, opinions and questionnaire responses were valid. That is they reflected accurately the behavior they were purported to be judging. However verbal behavior (i.e., writing, speaking, thinking) is controlled by the same behavioral principles (e.g., reinforcement, punishment, extinction) as non verbal behavior. Different meanings and values are differentially reinforced in the same manner as one teaches a child using differential reinforcement to discriminate between the colors red and blue. Albeit the differential reinforcement used to teach meanings and values may not be as overt as that used to teach color discrimination, it is nonetheless just as effective. Therefore self reports of behavior (as in questionnaires and interviews) are a function of the contingencies that control that particular behavior. Those contingencies may not be the same as (or have anything to do with) those which control the behavior they report on. The behavioral objection with the use of questionnaires, interviews and opinions is not that they are subjective, but whether or not they are accurate (i.e., true).

Both Geller (1991) and Hawkins (1991) have called for a functional approach to the use of methods of social validation.
In discussing the use of safety belt interventions Geller (1991) commented "the use of vehicle safety belts can only be considered socially valid if the relevant target groups use the intervention techniques when made available" (p. 180). The use of behavioral interventions beyond the conditions of an experiment is a problem of generality and as Stokes and Baer (1977) observed generality must be programmed in order to be maintained by the ongoing conditions in the environment. These conditions are not necessarily those used by the experimenter in the context of the experiment to achieve experimental control.

Social validity has also become synonymous with the terms "consumer satisfaction" (Hawkins, 1991) and "acceptance" (Kazdin, 1980). These terms are used in the sense that they are "predictors" of consumer (i.e., teacher or student) behavior at a time when they are still under the control or in contact with the contingencies of the experiment. Hawkins (1991) has proposed seven valid reasons for the use of measures of consumer satisfaction. Three of Hawkins' proposals are particularly relevant to this study:

1. To suggest promising adjustments to the procedures.
2. To predict or detect undesired effects.
3. To assess the comprehensiveness of the effects.

Such a view is consistent with an ecological (interactive) approach to experimentation. A functional assessment of the "consumer satisfaction and the acceptance of the procedures" may be obtained if teachers were to "use" the interventions in
other classes not under experimental control, or if teachers behaved in a manner consistent with their self reports in interviews, or if students "choose" from a menu of treatments, the treatment they preferred to operate in their classes. Accordingly, the tactics used in this investigation to assess "consumer satisfaction and acceptance of the procedures" were included in the experimental questions:

1. In a forced choice between treatments which treatment is selected by the students? (Experimental question 1.3). This question was determined by the arrangement of an experimental design to enable students in each class to make a choice (as to which treatment they preferred to operate in their classes) between all experimental conditions including the baseline conditions.

2. Which (if any) of the treatments are used by the teachers in their other classes? (Experimental question 1.4). This question was determined through the use of anecdotal records throughout the study provided by informal observation of other classes taught by the teachers.

3. Berliner (1986) suggested in his study of expert teachers that differential treatment by the teacher of students in the class can frequently be explained in terms of the history of the teacher and the student in the particular setting. Since this history is inaccessible (it is past) techniques to understand the nature of this past interaction (e.g., knowledge of the student) which influence current conditions have been developed.
Berliner (1986) used “stimulated recall” and “think aloud” techniques in his study of teacher planning. In their treatment of behavioral disorders Rolider and Van Houten (1985) “recreate the situation.” Since it is an assumption that an ecological view requires a micro-analytic analysis and since ignoring the history of interactions between teachers and students would be incomplete and contextually “thin” this study employed the following tactic to assess the teacher’s knowledge of the student. The procedure was not limited to teacher-student history alone, it also provided an opportunity for teachers to predict or detect undesired effects, and to suggest adjustments to the procedures (i.e., treatments). Replaying the videotape to present to the teacher critical vignettes defined either by the researcher’s inability to define a task congruency, or because of some contextual significance. Prior to the playing of a vignette the researcher posed one or some of the following questions

1. Can you help me to understand what went on here?
2. Why did you do what you did? and /or
3. Is there anything here I should be aware of?

The answers to questions 1 and 3 served the ecological function between researcher and teacher that Hawkins (1991) recognized. Question 2 provided some answers to the history of the teacher and student and was supported by observing other interactions between the teacher and this student to see if similar and consistent conduct is demonstrated by the teacher.
No attempt was made to prevent teachers from discussing the procedures or effects of the study with each other. In fact teacher interaction is viewed by the investigator as highly favorable. Teachers were however, kept naive to the dependent variables of the investigation.

Summary

This chapter described the methods of data collection involving machine observation and recording. Included were descriptions of the setting and the subjects, procedures describing the conduct of the machine observation and recording (videotaping) and a discussion of the strategy designed to reduce subject reactivity.

Next, the transduction process involving human observation and recording was presented. The response classes to be investigated were defined, and contingency relations were framed in terms of the antecedents defined as lesson component, task type and task specification, the response (task topography), and consequences which were described as response products and teacher behaviors. The instrument used to record the data was examined and the procedures used to measure interobserver agreement of the dependent variables and the procedural reliability of the independent variables were explained.

The third section of the chapter introduced the three treatments used in this study and described the arrangement of the instructional conditions for each treatment. Next the
methods of data analysis were examined and the research design and the tactics of graphic display of the data were elucidated. The chapter concluded with a discussion of social validity and the strategy used to measure consumer satisfaction was presented.
CHAPTER IV
DATA PRESENTATION AND DISCUSSION

This chapter presents the results and discusses the findings of this investigation. The investigation consists of three studies (basketball, volleyball, & badminton) in which three treatments (hustles, sprints and peer sprints) were implemented. Each successive study (i.e., volleyball and badminton) served to systematically replicate the treatments in a different setting. The treatments were applied to two classes Ms. East, and Ms. West with four target students selected for observation per class (n = 8 students). Thus by systematically replicating the treatments used in Class 1, the second class provided additional verification of the treatment effects. The same target students were used in all studies.

Accordingly, data are arranged by study and presented relative to the target students in each class. The experimental questions (relative to the entire investigation) will be discussed following the presentation and discussion of data. Similarly, social validity for the investigation will then be discussed.
Interobserver Agreement

Dependent Variables

Data were collected from a videotape record where a running digital time readout was imprinted on the videotape. This allowed the researcher to record three characteristics of the dependent variables: the temporal locus (i.e., when in time), the temporal extent (for how long), and repeatability (how many times an event occurred). Data on the dependent variables includes measures of the lesson component, the task type (e.g., refining or extending), the task specification (explicitness), student responses (a combined typographical and product judgment), teacher behavior and the use of time.

Interobserver agreement was conducted on 30% of all dependent measures collected. Random selection of videotapes were made from each class. Table 6 presents the percent agreement for the each of the dependent variables. The percent agreement for each dependent measure are discussed below.

Lesson Component, Task Specification, & Task Type

Data were kept on the amount of time spent in each of the lesson components. The lesson components were broadly categorized into three response classes commonly reported in the time-based literature. These classes were general content (time spent on lesson related tasks other than instruction and activity such as wait, transition, warm up and management),
Table 6

Interobserver Agreement Measures for the Dependent Variables

<table>
<thead>
<tr>
<th>Categories</th>
<th>Basketball Study</th>
<th>Volleyball Study</th>
<th>Badminton Study</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson Component</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Task Specification</td>
<td>96.6%</td>
<td>96.3%</td>
<td>100%</td>
<td>97.6%</td>
</tr>
<tr>
<td>Task Type</td>
<td>96.6%</td>
<td>100%</td>
<td>100%</td>
<td>98.9%</td>
</tr>
<tr>
<td>Student Responses</td>
<td>93.3%</td>
<td>95.4%</td>
<td>92%</td>
<td>93.6%</td>
</tr>
<tr>
<td>Teacher Behavior</td>
<td>91.4%</td>
<td>89%</td>
<td>93%</td>
<td>91.1%</td>
</tr>
</tbody>
</table>

Mean 96.3% 96.8% 97.5%

Note: All IOA data are reported as a total percentage per category.

knowledge (time spent by students on listening to instructions), and activity (time spent by students engaged in activity). Using the digital time recording from the videotape record observers verified the amount of time the class as a whole spent in each category. The percent agreement was 100% given a criterion of plus or minus 5 seconds (i.e., agreement occurred if the duration of the response class coded by the investigator occurred within
plus or minus of 5 seconds of the same event coded by a second observer).

Interobserver agreement for the dependent measures of task specification, and task type were treated as event data. The mean percent agreement for the three studies for task specification, and task type reported in Table 6 were 98.9%, and 97.6% respectively. These percentages were easy to determine and verify from the videotape record.

**Student Responses**

Data on student responses were treated as interval data. All responses that occurred during a treatment (and its baseline) were treated as occurring in intervals (measurement periods converted to rate per minute). Though the data paths are representative of a single response rate in fact they are the mean response rates for the behaviors under investigation. For example, 1-minute timings, an instructional arrangement of the sprint and peer sprint treatments were repeated as much as 6 times per task during any one day as a natural part of the instructional organization of the lesson. These 1-minute timings were averaged to produce a mean rate of response for the treatment.

Initially attempts were made to record both the topography and the response product. However capturing both the topography and response product for all four target students proved
unworkable using a single video camera. Rather than capturing an incomplete record, only measures on topography were used since these were the focus of the teachers' contingencies and instruction. Thus references in this chapter to correct or incorrect (error) rates and total response rates refer to topographical measures only. The mean percent agreement for student responses shown in Table 6 was 93.6%. Agreement was highest during the volleyball study (95.4%) where response classes were easily observed and discriminated, and lowest for the badminton study (92%) where the responses were more difficult to discriminate due to the location of the camera and the speed of a response.

Teacher Behavior

The mean percent agreement for teacher behavior was 91.1%. Agreement on teacher behavior was confounded by judgments by each observer regarding the purpose of the instructional event. Such judgments were made exclusively from the video tape record, and therefore were not confirmed by the teacher.

Independent Variables

Checks on the fidelity of the three treatments were conducted on 30% of all data collected and distributed equally among each treatment. Procedural fidelity was assessed in three ways. Where the treatment was time dependent (e.g., a 1-
minute timing) the videotape record with its imprinted digital time record provided verification that the duration of the treatment was consistent with the treatment description described in Chapter 3. There were 6 occasions when timings exceeded the 5 second criterion used to determine the fidelity of the timing (there were in excess of 500 one-minute timings during this investigation). In these cases data were counted only for the specified duration, responses occurring after the 60th second were ignored.

A second measure of procedural fidelity was kept using a procedural checklist (see Appendix C). The absence or presence of each item on the checklist was verified for each treatment. The percent occurrence for each teacher was 100% indicating that all treatments were faithfully implemented, as described in Chapter 3. Interobserver agreement was 100% for this checklist.

The hustles treatment required teachers to produce a rate of prompts and hustles that equaled or exceeded baseline rates fivefold. There were 6 occasions when Ms. East implemented the hustles condition with her class. Sixty-seven percent of the time Ms. East met criterion (4 of 6 occasions). Similarly, Ms. West implemented the hustles condition with her class 6 times, but met criterion only 17% of the time (1 of 6 occasions). Ms. West delivered a mean rate of prompts and hustles approximately threefold above baseline, well below the criterion
of fivefold over baseline. The interobserver agreement on the combined teacher rates was 86%.

A Procedural Note

Though the experimental questions have guided the design arrangements in this study, the investigator has been led by the data throughout. At times questions arose that occasioned changes and new design arrangements in the next study, as in the case of the volleyball and the badminton studies where across task comparisons were added. At times being led by the data has meant abandoning a treatment because teachers felt that it was inappropriate to pursue (as was the case with the hustle treatment in the volleyball study). Being led by the data has also occasioned changes during the conduct of a study which occurred in the basketball study when baseline and hustle treatments were applied to an additional task (passing) at the teachers’ request. None of these changes however, compromised the procedural fidelity or the integrity of the dependent variables that were collected. Though such changes are uncommon, they are not inappropriate when one uses an inductive strategy in scientific inquiry.
The Basketball Study

Framing the Study

This, the first of three studies sought, to ask the general question of what can be done to boost intensity (defined as rate of response) and what effects did this boost have on other variables, namely the qualitative dimensions of the response (correct technical performance i.e., topography); time use and task statements.

Data Presentation

Three sets of data will be displayed and discussed in this study. Data are presented concerning how students in the class spent their time (Figure 9), data reporting on the type and specification of the task (Table 8), and an alternating treatment design presenting the effects of the treatments (Figure 10).

Class 1: Results

Four students were selected for observation in this study. In accord with the criteria described in Chapter 3, students were selected on the basis of their gender and their skill level. Table 7 displays the relevant background information on each student reporting a pseudonym for each student, the student's skill level and gender.
Table 7
Background Information of Target Students in Class 1.

<table>
<thead>
<tr>
<th>Target Student</th>
<th>Skill Level</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeff</td>
<td>High</td>
<td>Male</td>
</tr>
<tr>
<td>Ian</td>
<td>Low</td>
<td>Male</td>
</tr>
<tr>
<td>Susan</td>
<td>High</td>
<td>Female</td>
</tr>
<tr>
<td>Beth</td>
<td>Low</td>
<td>Female</td>
</tr>
</tbody>
</table>

The basketball unit was taught for 12 days. Data were collected for 10 of these days. Of the remaining days, 1 was a multicultural day which resulted in a canceled class, and on the other day students received a written test on basketball rules. Thus data were collected on all days during which activity occurred. Figure 9 presents the percentage of class time spent on lesson content. General content ranged from 27% (Day 9) to 59% (Day 5) with a mean of 39%. Knowledge (mean 26%) ranged from 13% (Day 5) to 41% (Day 1). While time spent on activity ranged from 28% (Day 2) to 46% (Day 7) with a mean of 35%. All data paths display variability with three distinct patterns. During Days 1 and 2, knowledge was allocated the most time (approximately 40%) of the three lesson content areas with general content and activity each sharing the remaining 60% between them fairly equally. During Days 3 through 6, a second
<table>
<thead>
<tr>
<th>Task</th>
<th>Explicit / Code</th>
<th>Session / Code</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dribble to baseline RH &amp; LH</td>
<td>FX</td>
<td>E1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing Crossover</td>
<td>FX</td>
<td>E</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running Crossover</td>
<td>FX</td>
<td>E2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Break Jump</td>
<td>FX</td>
<td>E</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breakdown</td>
<td>FX</td>
<td>E3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run jump stop no ball used</td>
<td>FX</td>
<td>E</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shooting in pairs</td>
<td>FX</td>
<td>I</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hustle (I₁)</td>
<td>FX</td>
<td>H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sprint Condition C (I₁, E₂, &amp; E₃)</td>
<td>FX</td>
<td>C</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sprint Condition D (I₁, E₂, &amp; E₃)</td>
<td>FX</td>
<td>D</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shooting Baskets</td>
<td>FX</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choice 3x3, knockout etc.</td>
<td>PX</td>
<td>A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chest Pass</td>
<td>FX</td>
<td>I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bounce Pass</td>
<td>FX</td>
<td>I₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hustle Bounce Pass</td>
<td>FX</td>
<td>H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defensive Stances</td>
<td>PX</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Informing (I); Extending (E); Refining (R), Applying (A), Sprint (C), Peer Sprint (D), Hustle (H), Fully explicit (FX), Partially explicit (PX), & Implicit (IM). X indicates frequency of task occurrence. I₁, E₂, E₃, I₂, indicate baseline drills used in interventions.
Figure 9. Percentage of class time spent on lesson content for Class 1 during the basketball study.
Figure 10. Correct responses per minute for each target student in Class 1.
pattern emerges. During this time general content accounted for 40% to 59% of the lesson time, activity approximately 30% and knowledge approximately 20%. During days 7 through 10 knowledge remained at 20% (with a slight upward trend), general content was approximately 35% and time spent engaged in activity occupied around 40% of the lesson (with a slight downward trend).

Table 8 displays the nature, explicitness, type and frequency of Ms. East's task statements during this study. The most common task statement was an extending task (n = 7) followed by informing tasks (n = 5). There were no refining tasks coded, however there was a single applying task. All but 2 tasks (which were partially explicit) were fully explicit. The interventions (identified in Table 8 as C, D, & H) were the most frequently recurring tasks.

The baseline, a mean of three tasks (identified in Table 8 as I1, E2, E3, & I2) and the three treatments (hustles, sprints & peer sprints) were compared using an alternating treatments design. A response was a single bounce of the ball performed in the context of the task (e.g., dribble to the mid-line keeping the ball bouncing at hip height, changing hands at designated markers). Figure 10 presents this design and the rates of correct responding for each of 4 target students (2 high skilled and 2 low skilled students with each gender represented at each skill level). The difference between correct responses and total
responses was negligible. Thus for the purposes of these data it was assumed all responses were in fact correct responses (data paths actually present the more conservative measure of correct response data). Additionally, data for the 4 target students was similar enough that trends and levels can be described jointly.

Table 9 displays the means and ranges for correct responses for baseline and treatment of dribbling and passing. Baseline data were collected throughout the study on two response classes, dribbling tasks (first 6 tasks listed in Table 8) and passing tasks (chest and bounce pass).

All treatments focusing on dribbling produced rates that exceeded 35 responses per minute (except for Ian on Day 3, with 30 rpms). Baseline data for dribbling tasks displayed in Figure 10 indicate an initial peak (the highest was Jeff at 36 responses per minute or “rpm”) after which data paths descended on Day 3 to below 28 rpm for all students, and remained both stable and level for the remainder of the study. Baseline data on passing reveals rates that did not exceed 10 rpm for any student. The hustle treatment was applied to two separate tasks, dribbling and chest passing. Hustle rates obtained for dribbling for all students exceeded 35 rpm (the highest obtained by Susan at 68 rpm). Hustle rates for chest passing ranged between 5 and 10 rpm for all students (Susan & Beth were lowest). Sprints produced rates above 40 rpm for all students and as high as 53 rpm (Jeff, Ian & Susan on Day 3). Peer sprints remained stable,
Table 9

Means and Ranges of Correct Responses for each Task Across Treatments for each Target Student in Class 1.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Jeff</th>
<th>Ian</th>
<th>Susan</th>
<th>Beth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline/</td>
<td>21</td>
<td>24</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Dribbling</td>
<td>17-36</td>
<td>15-29</td>
<td>20-25</td>
<td>15-32</td>
</tr>
<tr>
<td>Hustle/</td>
<td>42</td>
<td>46</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>Dribbling</td>
<td>37-56</td>
<td>31-53</td>
<td>31-53</td>
<td>34-51</td>
</tr>
<tr>
<td>Sprint</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>46-53</td>
<td>45-54</td>
<td>45-53</td>
<td>41-52</td>
</tr>
<tr>
<td>Peer Sprint</td>
<td>48</td>
<td>51</td>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>45-52</td>
<td>48-57</td>
<td>48-57</td>
<td>42-52</td>
</tr>
<tr>
<td>Baseline/</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Passing</td>
<td>4-8</td>
<td>4-7</td>
<td>4-7</td>
<td>3-4</td>
</tr>
<tr>
<td>Hustle/</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Passing</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>6-7</td>
</tr>
</tbody>
</table>

* signifies a range of zero

and high throughout the study with the modal rate occurring between 45 to 49 rpm for each student. The highest rate of 57 rpm was obtained by Ian on Day 9.
Class 1: Discussion

Time Use

Data displayed in Figure 9 indicate that considerable time was spent throughout the study in general content. General content included content (wait & transition) and non-content related (attendance) activities and while it appears to be a higher percentage of the lesson, it represents a cumulative percentage of several categories (see Chapter 3). Despite the cumulative effect of various subcategories much of the general content percentage was explained by the time taken for attendance. Taking attendance typically took between 5-10 minutes daily. Initially knowledge had the highest percentage of the three content areas which suggests the significance of instruction (i.e., informing and instructing) during the first days of this unit. Following Day 2 knowledge consisted of stating tasks and explaining the treatments. Surprisingly, general content rose significantly during the middle of the study. Perhaps this may be explained in terms of the new organizational demands brought about by the treatments (pairs of students, handing out peer coding cards). Activity had the highest percentage of the three content areas during the last third of the study. In part, this can be explained in terms of the “routinization” of the treatments (they became increasingly more time efficient to set up and run), and in part, because the teacher allocated more time to activity, particularly to tasks
where students could choose the specific activity they wished to pursue.

**Tasks**

The frequency with which tasks recurred over the duration of the unit (see Table 8), indicates that with the exception of the treatments (C, D, & H) and the applying tasks, few other tasks were used on more than 2 days. The single applying task consisted of a choice from among three or more small groups of 3 on 3 activities such as round the world, and half court games. During these activities students were generally equally distributed among activities.

Data displayed in Table 8 presents a picture of skill development where game play, other than in the choice option was noticeably absent. Ms. East deliberately did not plan for games to be included in her lesson. When asked why, Ms. East stated that playing games of basketball “would be unfair” because of the wide range of skill level in most of her classes. The issue of not leveling the playing field (i.e., creating equitable opportunities for game play across gender and skill levels) was not seen as problematic by Ms. East. Though Ms. East did see the inequity of the situation as problematic, she choose to maintain the focus on skill development and avoid playing games. It may be that this was Ms. East’s effort to correct skill differences. However, as the data to be reported in the
volleyball and badminton studies will show, the skill level was equally varied in these sports, yet the majority of the activity time was spent on game play.

Though officially an invasive, but non-contact sport, basketball is frequently fraught with physical contact. Under these conditions, physical size and aggressiveness become factors not only from a safety standpoint, but from a tactical one as well. One explanation for Ms. East's reticence about engaging in game play during basketball might be the issue of safety. Neither volleyball nor badminton (both net games) involve physical contact with other teams, and thus safety from physical contact is not an issue (there are however several other safety issues associated with each of these sports unrelated to physical contact, such as falls to the floor in volleyball and shuttlecock-eye damage in badminton).

From observations, presenting knowledge was an important part of Ms. East's job. She took time to question and to discuss the tasks with her students. Tasks were explicitly stated most of the time and were restated through the course of the unit. Typically Ms. East did not provide the students with information, rather she questioned them and asks them to discriminate her behavior. For example, in discussing the jump shot she might hold the ball a particular way and ask the question "Is this right?" then request of a student "Come over and show me how I should hold it" followed by "Is this right?" Students
participated in this style of questioning and though it took time, students did correctly discriminate her behavior. These discriminations served as the criteria for not only baseline measures of responding, but also for the intervention. This feature of Ms. East's instruction was particularly noticeable when she reviewed the treatments each day, and it appears to be a part of her day to day instruction. During informal observations of her other basketball classes which were not apart of this investigation, the explicitness of Ms. East's task statements and her organization of tasks were entirely consistent with the description provided in Table 8. This suggests that Ms. East has "standardized" her approach to instruction across classes.

General Treatment Effects

All treatments focusing on dribbling produced effects (increases in rates) at least twofold over baseline rates. All treatments (hustle, sprint & peer sprint) produced this effect in dribbling responses (hustles used in passing will be discussed separately). For example, Ian who had the highest mean at 24 rpm had increases in magnitude to 48 rpm (approximate mean for all treatments). Baseline dribbling tasks differed from treatment tasks in that (a) they were not timed by the teacher (baseline times were varied usually greater than 3 minutes, but less than 5 minutes throughout the duration of the study) and (b)
students frequently walked or rested during the tasks (probably related to the duration of the task). The baseline tasks were similar in that it was the same task, with the same performance criterion (excluding timing), and was conducted in the same pairs.

**Hustles**

The hustles treatment displayed the most variability. Though initially producing the highest or near highest levels of responding for all students its effect on subsequent days produced high, but variable responding for Ian, Susan, and Beth. Jeff’s responding during hustles remained stable following Day 1 at 36 rpm. Ian benefited most from the hustles. He consistently produced rpm’s above 50. For Susan hustles produced high levels of responding throughout the intervention particularly on Day 1. Beth the least skilled of the four students responded well to the hustles treatment. Her data indicate that she was increasing her level of responding as time progressed.

Day 1 produced high levels of responding for all students during hustles. Perhaps the novelty of a “vocally hyperactive” teacher is an explanation. Hustles usually lasted for approximately 5 minutes. The longevity of the effectiveness of this treatment is unknown, since the treatment was used 4 times during dribbling and twice during passing, and for only 5 minutes or so at any one time. Procedural fidelity was weakest
during hustles (Ms. East was only faithful 68% of the time to the fivefold criterion) because Ms. East found that keeping up a continuous hustle pace (e.g., "go, go go...come on...Susan,...Well done Beth") was physically demanding (among other things voice strain was a problem). It was not as Ms. East put it "something I could do every day or even for short periods in each lesson." In addition, Ms. East felt that it focused her attention on one component of her instruction at the expense of others. In short, the response cost for Ms. East was too high.

Hustles were also used to intervene upon passing. The goal here was to see what effect hustles would have on an activity like passing which by its nature did not produce high response rates, like those found in dribbling tasks, but which relative to the task can still be high. In dribbling, the ball is under the control of the performer, and a response can be made at close to 1 per second. The passing task arranged by Ms. East (the first person in a line of 3 passed to the first in another line of three) could not generate such rates as 40 or 50 per minute (though cumulatively it might come close if the total passes per line were added up). The findings for Jeff, Ian, Susan and Beth (Days 7 & 8) indicate that hustles might be marginally better than baseline. There appeared to be some induction or generalization of treatments on Day 8 for Jeff, Ian, and Susan, but two data points are insufficient upon which to base a conclusion. Important questions occasioned by this "probe" are (a) "What if
the effects of high response rates are a product of the instructional grouping (pairs as opposed to lines) rather than the task," (b) "Are the effects on the level of responding relative to those activities that can generate a high response rate (e.g., dribbling versus golf shots)" and (c) "What role does task difficulty and/or task complexity play in correct responding?"

**Sprints and Peer Sprints**

The sprint and the peer sprint conditions were most effective. The treatments produced similar effects across gender and skill level and these effects persisted throughout the study producing consistently high levels of correct responding. It is hard to differentiate between the two in terms of effects. A case might be made on the basis of Ian and Susan that peer sprints were more effective. However there was considerable crossover (overlap) between treatments for Jeff and Beth to making this conclusion (peer sprints superior to sprints) speculative. It was certainly the investigators' and Ms. East's observations that peer sprints were more effective. But these observations are not based on data presented in Figure 9, but rather they are based on variables not measured in this study, but nonetheless observed. For example, there appeared to be no off-task behavior during the peer sprints. Students during peer sprints would adopt standing racing start position before the sprints began, and recorders would be ready to code correct and
incorrect responses. Students were “motivated” by the peer sprints and appeared eager and task directed throughout the study. No overt inter-student competitiveness or misbehavior (e.g., arguing over who was best etc.) was observed by the teachers or investigator.

Class 2: Results

Four students were selected for observation in this study. Table 10 displays the relevant background information on each student in Class 2 taught by Ms. West.

Table 10

<table>
<thead>
<tr>
<th>Target Student</th>
<th>Skill Level</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>High</td>
<td>Male</td>
</tr>
<tr>
<td>Tim</td>
<td>Low</td>
<td>Male</td>
</tr>
<tr>
<td>Carol</td>
<td>High</td>
<td>Female</td>
</tr>
<tr>
<td>Linda</td>
<td>Low</td>
<td>Female</td>
</tr>
</tbody>
</table>

The basketball unit was taught for 12 days. Data were collected for 11 of these days (Ms. West’s class was unaffected by the multicultural day). On the remaining day a written test on basketball rules was given to the class. Data were collected on
all days during which activity occurred. Figure 11 presents the percentage of class time spent on lesson content. General content ranged from 24% (Day 6) to 59% (Day 1) with a mean of 44%. Knowledge (mean 17%) ranged from 6% (Day 10) to 26% (Day 4). Time spent on activity ranged from 25% (Day 8) to 57% (Day 11) with a mean of 40%. The data paths representing time spent on general content and activity are highly variable (almost cyclical) and display no pattern other than a crossover effect every other day. Time devoted to instruction (i.e., knowledge) remained low and relatively stable throughout the study.

Table 11 displays the nature, explicitness, type and frequency of Ms. West's task statements during this study. The most common task statement was an extending task (n = 6) followed by informing tasks (n = 3). There were no refining tasks coded and there was 1 applying task. Three tasks were partially explicit the remainder were fully explicit. The interventions (identified in Table 11 as C, D, & H) were the most frequently recurring tasks.

Figure 12 displays the data for John, Tim, Carol and Linda. Linda was absent on days 5 and 10 and was injured and unable to complete the lesson on day 8. Since the data for the 4 target students are similar, trends and levels have been described jointly. Table 12 displays the means and ranges for correct responses for baseline (the mean of tasks $l_1$, $E_2$, $E_3$, & $l_2$ in Table 11) and treatment of dribbling and passing. All of the
Table 11

Mrs. West's Task Statements during the Basketball Study.

<table>
<thead>
<tr>
<th>Task</th>
<th>Explicitness</th>
<th>Session Code</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dribble to baseline RH &amp; LH</td>
<td>FX</td>
<td>( I_{1} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Standing Crossover</td>
<td>PX</td>
<td>E</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running Crossover</td>
<td>FX</td>
<td>( E_{2} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breakdown</td>
<td>FX</td>
<td>( E_{3} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run jump stop no ball used</td>
<td>FX</td>
<td>E</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hustle (( I_{1} ))</td>
<td>FX</td>
<td>H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sprint Condition C (( I_{1}, E_{2}, &amp; E_{3} ))</td>
<td>FX</td>
<td>C</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sprint Condition D (( I_{1}, E_{2}, &amp; E_{3} ))</td>
<td>FX</td>
<td>D</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shooting Baskets</td>
<td>FX</td>
<td>E</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choice 3x3, knockout etc.</td>
<td>PX</td>
<td>A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shooting Baskets</td>
<td>PX</td>
<td>E</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chest Pass</td>
<td>FX</td>
<td>I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bounce Pass</td>
<td>FX</td>
<td>( I_{2} )</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hustle Bounce Pass</td>
<td>FX</td>
<td>H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Informing (I); Extending (E); Refining (R), Applying (A), Sprint (C), Peer Sprint (D), Hustle (H), Fully explicit (FX), Partially explicit (PX), & Implicit (IM). X indicates frequency of task occurrence. \( I_{1}, E_{2}, E_{3}, I_{2} \), indicate baseline drills used in interventions
Figure 11. Percentage of class time spent on lesson content for class 2 during the basketball study
Figure 12. Correct responses per minute for each target student in Class 2.
high rpms in the range occurred on Day 1. Baseline/dribbling data never exceeded 35 rpm for the duration of the study.

Table 12
Means and Ranges of Correct Responses for each Task Across Treatments for each Target Student in Class 1.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>John</th>
<th>Tim</th>
<th>Carol</th>
<th>Linda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline/Dribbling</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>18-35</td>
<td>19-32</td>
<td>19-35</td>
<td>19-24</td>
</tr>
<tr>
<td>Hustle/Dribbling</td>
<td>49</td>
<td>47</td>
<td>53</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>37-67</td>
<td>29-70</td>
<td>39-74</td>
<td>35-68</td>
</tr>
<tr>
<td>Sprint</td>
<td>47</td>
<td>54</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>39-57</td>
<td>47-59</td>
<td>43-58</td>
<td>43-55</td>
</tr>
<tr>
<td>Peer Sprint</td>
<td>56</td>
<td>56</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>52-68</td>
<td>48-60</td>
<td>48-57</td>
<td>47-51</td>
</tr>
<tr>
<td>Baseline/Passing</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>5-7</td>
<td>*</td>
<td>3-5</td>
</tr>
<tr>
<td>Hustle/Passing</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7-8</td>
<td>6-8</td>
<td>5-7</td>
<td>5-7</td>
</tr>
</tbody>
</table>

* signifies a range of zero
In general, the baseline data are similar and stable across students. The hustles/dribbling treatment produced considerable, but similar variability for all students. The pattern of the variability was such that it peaked on Day 3 (John 67, Tim 70, Carol 74, and Linda 68 rpm), then descended on Day 4 to rise again on Day 5. The means for the students for the hustles/dribbling treatment are John 48, Tim 47, Carol 53, and Linda 46 rpm.

The sprint condition produced a relatively level rate of responding across throughout the study for all students. John displayed the most variability during Days 10 and 11. Similar, but higher levels of correct responding were recorded for the students during the peer sprint condition.

**Class 2: Discussion**

**Time Use**

The variability in the general content and activity data for Class 2 has several potential causes. This class was the only physical education class that Ms. West taught alone. Two other classes were team taught with Ms. East. The remainder of the time Ms. West taught health. Ms. West was not as organized as Ms. East in her day-to-day organization. Two assistants who helped the investigator in the videotaping of these classes each remarked that Ms. West was not as organized as Ms. East. Ms. West seldom completed taking attendance in less than 12
minutes. During attendance she often left the gymnasium and returned briefly to her office in the girls locker room for a moment or two. Additionally, Ms. West spent at least 5 minutes every other day reprimanding the class in one manner or another for misconduct. This day-to-day inconsistency in Ms. West's teaching behavior may be an explanation for the variability in the general content data and thus the variability in available time for activity.

Tasks

The similarity of Ms. West's data to Ms. East's data requires explaining. A typical day for Ms. West was as follows. First period health was taught during the time Ms. East and her class were involved in the investigation. Following period 1 Ms. West team taught with Ms. East for periods 2 and 3. The grade levels differed but the content and instructional arrangements did not. As mentioned previously Ms. East (Class 1) often used a standardized approach to her instruction (same tasks, same sequencing of tasks, and frequently the same arrangement of time). Frequently the treatments were used and experimented with during periods 2 and 3 (refer to comments on social validity and generalization at the end of this chapter). The lessons for periods 2 and 3 served as models for the lesson Ms. West would teach in period 4 during which the intervention was conducted. In fact Ms. West commented that this helped her cope with a
half-time physical education and half-time health appointment. Thus it is not unreasonable to conclude that the similarity between Ms. West’s tasks and Ms. East’s tasks was a product of the team teaching experience which was mostly under the direction of Ms. East.

Ms. West used fewer extending tasks than Ms. East and more of her task statements were partially explicit. Like Ms. East, Ms. West did not include games in her plan other than providing for the option of a 3 on 3 game during the applying task. She like Ms. East did not find the issue of leveling an uneven playing field as problematic, or if she did she avoided it by not playing games.

**General Treatment Effects**

All treatments focusing on dribbling produced effects similar to those reported in Class 1. Increases in rates were at least twofold over the baseline rates. The largest gains were made by John and Tim. Their respective mean baseline rates of 24 rpm increased to 49 and 47 rpm during hustles, 47 and 54 rpm during sprints, to 56 rpm each during peer sprints.

**Hustles**

As was the case for Class 1 the hustles treatment was applied to 2 instructional tasks (dribbling and passing). The treatment hustles/dribbling reproduced the effects found in
Class 1. In Class 1 there were marginal gains from baseline to intervention when the hustles/passing treatment was used. The effects of this treatment on the students in Class 2 was negligible. In part this must be attributed to poor procedural fidelity (Ms. West faithfully implemented the treatment on 1 of 6 occasions). Yet Day 5 when the criterion was met, was not the day of the greatest rpm, which was Day 3 for all students. Data from the videotape record did not reveal why Day 2 should produce such high levels of responding, other than the frequent use of names more on Day 3, than on other days. A similar jump occurred in Class 1, but it was attributed to the novelty of a "hyperactive teacher." It is more difficult to use this explanation here. The fivefold increase over baseline was an arbitrary decision by the investigator. Perhaps it was not the frequency on hustles, but the "personalizing" of the hustle (letting students know exactly that they are, or they are not completing the task correctly). Ms. West echoed the feelings of Ms. East by questioning the value and the response cost (voice strain) of the hustles treatment.

Sprints and Peer Sprints

The effects here are clear. Both sprints and peer sprints produced significant improvement in correct responding for dribbling tasks. In Class 1 differentiating between treatments on the basis of response rate was considered speculative. In
Class 2 the results are more defined. There was minimal crossover of the data paths (see Tim and Carol, Day 4) and it occurred early in the interventions. The data paths for all students indicate that peer sprints are consistently more effective than sprints. However some cautionary comments are warranted. Though peer sprints were more powerful than sprints, 3 of the data paths of the students for sprints were ascending as time progressed. This might be the effect of induction/generalization, or it might be the effect of the improvement in the sprint condition alone. Future replications of these conditions may provide the answers. Like Class 1 (and more noticeably so) students in Class 2 were on task more than in the other conditions (hustles, baseline, sprints). They were “ready” for activity.

**Forced Choice Between Treatments for Class 1 and Class 2**

On the last day of the unit the teachers gave students a choice (majority rule) regarding which treatment they would like to use that day while they practiced the drills. In each class the decision was quickly arrived at and as Figures 10 and 12 indicate there was agreement between the two classes as to the preferred treatment. The sprint condition was overwhelmingly preferred by the majority of the students in each class. Some students in each class did opt for the peer sprints, none chose in favor of the hustles. Student comments captured on the
videotape record provide some insight into why sprints were preferred to hustles and peer sprints. There appeared general agreement in both classes that the students did not want hustles, however no interviews were conducted with the students to determine why. Comments made by students during lessons led the investigator to conclude that students saw peer sprints as having a higher response cost than sprints particularly in terms of the effort of observing and recording. Interestingly, enough, so to did the teachers.

Study 1 Conclusions

The conclusions relating to dribbling and passing are presented separately.

1. Both sprint and peer sprints are powerful treatments for the dribbling task chosen. In this study all students had their respective baseline responding doubled. It remains to be seen if they are as effective with other tasks where task complexity, or low response rates are a factor.

2. There was some evidence from Class 2 to suggest that peer sprints may be more effective than sprints.

3. The hustles/dribbling treatment produced variable, but high levels of responding. It did so at a cost. Teachers found that implementing this treatment was physically demanding and "instructionally myopic" (investigator's phrasing) focusing
attention on one aspect of instruction at the expense of others. Both teachers requested not to use it again.

4. The hustles/passing treatment was not an effective treatment. It was not significantly more effective than baseline. It was unclear if this was a product of an activity (passing) with a naturally low response rate per minute or if it was a product of the grouping (lines of 3's).

5. Neither gender nor skill level were variables which discriminated between performance in terms of effects (direction and significance) or level (magnitude of effect). This conclusion is relative to the nature of the task (basketball tasks were very easy) and dependent upon the accuracy of the teacher's initial judgment about student skill levels (were student correctly labeled low or high skilled?).

The Volleyball Study

Framing the Study

Based on the findings of the basketball study the volleyball study sought to replicate the treatments used in the basketball study with a view to reproducing the findings of the basketball study in a volleyball setting. Initially the investigator had intended to replicate the treatment hustles, however teachers did not want to continue this treatment (physically demanding), and expressed doubts regarding its effectiveness. Accordingly, the treatment was discontinued.
The design arrangement remained the same in this study, though the instructional arrangement differed from the basketball study. Two experimental treatments (sprints and peer sprints) were compared across 3 tasks to baseline measures. Whenever a treatment was implemented it was implemented on all of the tasks. The tasks were (a) bumping off a wall (wall drill), (b) having a volleyball lobbed (fed drill) to a partner who set it back, and (c) a 1 on 1 (1X1), set and bump interchange where partners bump and set back and forth with the goal of a continuous (non-interrupted) play. These tasks were commonly used by the teachers in their past and ongoing lessons. Like the basketball tasks, the students were paired together. The tasks differed by difficulty. The fed drill was seen as easier by the teachers than the wall drill, which was easier than the 1 X 1 drill.

The basketball study produced minimal error rates, in part because the tasks selected were relatively easy to perform for all students. Both Ms. East and Ms. West felt confident that this would not be the case in volleyball. They expected as wide a range of ability in volleyball as they had in basketball, however the drills chosen by the teachers to be intervened upon had in the past according to the teachers discriminated the students by ability. Thus in addition to rates of correct responding, rates of incorrect responding were included. Total levels of responding
have not been separately reported for the sake of economy. Total rates of responding may be obtained by adding the success and error rates together.

In the pre-study discussion the teachers informed the investigator that unlike basketball, volleyball would consist of organized game play nearly every day. The investigator saw this as an opportunity to collect data on responses during game play. No intervention occurred (i.e., treatments not used, descriptive focus only). The assumption of equitable opportunities was tested. For example in a 3 X 3 volleyball game physical educators might expect that responses might be distributed 33% to each player on a team. Though the outcome of a response might be determined by the skill of the student, the opportunity to respond should not. In this study 3 X 3, 4 X 4, 5 X 5, and 6 X 6 games were observed and data for each target student were reported as a percentage of the total responses made by the team in which the student was a member. Total responses made by the team were also reported. Rate was not reported, since rate was not a good indicator of an individuals response as a member of a team. The choice of the arrangement was left up to the teacher and at times was determined by the number of students in the class, rather than by a particular order. Though not directly intervened upon, team size did represent a “naturally” occurring intervention in the context of the lesson. Responses in this study are not the same as opportunities, there were
approximately 3-5 % more opportunities than responses because there were occurrences when either a response could not be made, or was not made.

Class 1: Results

The volleyball unit was taught for 12 days. Data were collected for 11 of these days. On the remaining days students received a written test on volleyball rules. Some of the target students were absent during this unit. Jeff was absent on Day 11, Ian on Day 3, and Beth on Days 8 & 11.

Figure 13 displays the percentage of class time spent on lesson content for Class 1. Data for each category are variable. The mean for general content was 39% (range 29%-52%). The knowledge category displays the most variability with a range from 13% to 56% (mean 26%). Data ranged between 15% and 47% for activity with a mean of 35%.

Three tasks (1 informing, 1 extending and 1 refining), all concerned with the serve, were used by Ms. East. These were the only additional tasks used in the unit other than games and the tasks used in the treatments. Seventy - three percent of tasks were fully explicit while the remainder (17%) were partially explicit. Tasks involving the treatments were usually fully explicit, while tasks involving games were mostly partially explicit.
Figure 14 presents a direct comparison between the treatments (excluding baseline) measuring correct responses per minute for target students across tasks. The means and ranges of correct responses across treatments and tasks are presented in Table 13. The means for peer sprints exceeded (equaled once see Susan Table 13) the means obtained for the sprint condition.

Table 13
Means and Ranges of Correct Responses for each Task Across Treatments for each Target Student in Class 1.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Jeff</th>
<th>Ian</th>
<th>Susan</th>
<th>Beth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wall Drill:</td>
<td>59</td>
<td>61</td>
<td>52</td>
<td>16</td>
</tr>
<tr>
<td>Sprint</td>
<td>50-75</td>
<td>60-70</td>
<td>45-55</td>
<td>10-24</td>
</tr>
<tr>
<td>Wall Drill:</td>
<td>57</td>
<td>69</td>
<td>52</td>
<td>29</td>
</tr>
<tr>
<td>Peer Sprint</td>
<td>51-64</td>
<td>61-73</td>
<td>45-58</td>
<td>24-32</td>
</tr>
<tr>
<td>Fed Drill:</td>
<td>11</td>
<td>14</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Sprint</td>
<td>9-12</td>
<td>12-15</td>
<td>11-13</td>
<td>4-6</td>
</tr>
<tr>
<td>Fed Drill: Peer</td>
<td>20</td>
<td>19</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Sprint</td>
<td>19-20</td>
<td>17-21</td>
<td>12-19</td>
<td>11-15</td>
</tr>
<tr>
<td>1 V 1 Drill:</td>
<td>11</td>
<td>15</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Sprint</td>
<td>7-14</td>
<td>8-28</td>
<td>8-16</td>
<td>1-4</td>
</tr>
<tr>
<td>1 V 1 Drill:</td>
<td>18</td>
<td>17.5</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Peer Sprint</td>
<td>18-19</td>
<td>12-20</td>
<td>13-20</td>
<td>7-9</td>
</tr>
</tbody>
</table>
With the exception of Beth (sprint mean 16/ peer sprint mean 29) all other means were above 50 rpm. Differences existed for all students between the means (both peer sprint and sprint) for the wall drill and the means for the fed and 1 V 1 drills. In general the wall drill produced higher rpm than did the fed or 1 V 1 which produced lower but similar findings (below 20 rpm).

Figures 15-17 present within-task comparisons of baseline and treatments using correct and incorrect responses for the wall drill (Figure 15), the fed drill (Figure 16) and the 1 V 1 drill (Figure 17). In Figures 15 and 17 baseline and treatments began on Day 4 due to the instructional arrangement of the unit by the teachers. In Figures 16 (the fed drill) baseline began on Day 1. The general pattern in Figures 15 through 17 was that peer sprints more consistently produced higher rates of correct responding than did the sprint condition and that incorrect responding illustrated in Figures 15 through 17 remained low during the treatments.

Data relative to the effects of team size on the proportion of responses made by a target student are displayed in Figure 18 (histogram display). The proportion of team responses expected relative to team size is indicated by the dashed line and identified by the arrow/percentage. Student share of the responses varies throughout the study as a function of team size. Beth received the least equitable share of responses, she consistently fell below
Figure 13. Percentage of class time spent on lesson content for Class 1 during the volleyball study.
Figure 14. Correct responses per minute for target students in Class 1 for Sprints and Peers Sprints across instructional tasks.
Figure 15. Correct and incorrect responses for the wall drill for target students in Class 1 during the volleyball study.
Figure 16. Correct and incorrect responses for the fed drill for target students in Class 1 during the volleyball study.
Figure 17. Correct and incorrect responses for the 1 V 1 drill for target students in Class 1 during the volleyball study.
Figure 18. Percentage of team responses made by target students in Class 1 (left y axis - histogram) and number of team responses (right y axis - freq. polygon) for students during the volleyball study. Proportion of team responses expected for the team size is indicated by horizontal dashed lines.
criterion. The total number of responses made by the team is represented by the frequency polygon (solid line). Team responses are variable, but the trend was ascending for all students. Games were not played everyday (see Days 1, 3, and 5). Some students were absent from game play (Jeff Day 11, Beth Days 10 and 11). The means of incorrect responses for each task across baseline and treatment conditions are presented in Table 14.

Table 14
Means of Incorrect Responses for each Task Across Baseline and Treatment Conditions for each Target Student in Class 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>Conditions</th>
<th>Jeff</th>
<th>Ian</th>
<th>Susan</th>
<th>Beth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wall Drill</td>
<td>Baseline</td>
<td>2.5</td>
<td>4.3</td>
<td>3.0</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>8.0</td>
<td>4.0</td>
<td>2.0</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>5.3</td>
<td>3.3</td>
<td>5.2</td>
<td>14.0</td>
</tr>
<tr>
<td>Fed Drill</td>
<td>Baseline</td>
<td>1.3</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>3.3</td>
<td>1.7</td>
<td>1.0</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>2.3</td>
<td>2.0</td>
<td>1.0</td>
<td>1.7</td>
</tr>
<tr>
<td>1 V 1 Drill</td>
<td>Baseline</td>
<td>1.0</td>
<td>0.5</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>3.0</td>
<td>2.5</td>
<td>1.5</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>3.0</td>
<td>3.0</td>
<td>2.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Note: All tasks of equal duration, and treatments applied equally to each task.
Class 1: Discussion

Time Use and Tasks

The arrangement of the lesson time in Study 2 was similar to Study 1. Though more time each day (from Day 4 onwards) was spent engaged in game play (a third to half of the activity time) the general picture of around 40% of the lesson spent in general content activities, and 25% spent on instruction (knowledge) and the remainder (35%) spent on activity is fairly consistent across days. The data path representing the knowledge category descended after Day 6 from 30% of the lesson to approximately 15% of the lesson for the remaining 5 days. This reflected the decreased role of instruction and the increased emphasis on game play during this time. The reverse occurred on Day 1 when knowledge accounted for nearly 60% of the lesson. Though the mean activity time data was similar for Class 1 across studies 1 and 2 (35% each), the use of that time was not. In Study 1, the majority of the activity time was spent on skill development related tasks. In study 2, for 8 of the 11 days game play occupied approximately 40% of the time allocated for activity. This shift of focus as previously discussed in Study 1 was a significant change in the instruction used to teach volleyball and basketball, both of which are team games.

There were only 3 additional tasks in this study (not including the tasks intervened upon). These were recurring tasks
for the first 3 days, and were devoted to the volleyball serve (informing, extending and refining). The majority of tasks were explicitly stated, though many game-tasks were partially explicit (usually a performance criterion was absent).

Direct Comparisons Between Treatments Across Tasks

Using correct responses per minute the data shown in Figure 14 for Jeff, Ian, Susan, and Beth are similar. Though the data for Beth are considerably less in magnitude than the data reported for Jeff, Ian, and Susan, the general effects of the treatments across tasks may be stated as:

1. Both treatments applied to the wall drill produced high levels of responding (above 40 rpm for Jeff, Ian, and Susan). The fed drill and the 1 v 1 drill produced levels of responding that did not exceed 25 rpm for either treatment.

2. In general, the effects of peer sprints produce responding that was higher than the sprint treatment for all students (though there was some crossover for Ian and Susan), more stable than the sprint treatment and for three of the students (Jeff, Ian and Susan) peer sprints maintained an upward trend.

Note: Figure 15 was scaled from 0 to 80 responses per minute to reflect the high levels of responding, similarly Figures 16 & 17 are scaled from 0 to 30 responses per minute to reflect lower levels of responding.
Wall Drill

Incorrect responses (i.e., error rates) are displayed in Figure 15 (as well as Figures 16 & 17) as open squares, circles, and triangles, correct responses are represented by closed symbols. Baseline correct responses for each student show low, but varied levels of responding (too few data points to label these data tends). Mean rates for incorrect responses across students in the wall drill are presented in Table 14. The mean incorrect responses for baseline for students were lower than either treatment means (exceptions were lan-sprints & peer sprints, & Susan-sprints). Incorrect responses had similar proportional increases to the gains made by correct responses over baseline. A general conclusion was that while increased responding produced a greater number incorrect responses, when responses are compared proportionally, the treatments did not increase the level of incorrect responses.

The wall drill produced high levels of responding for all students. Gains from baseline to treatments (see Table 13) indicate a doubling of responses. These gains are consistent with those reported for basketball (see Study 1). The wall drill in analogous to the dribbling in basketball. In both tasks the student controls the pace of the activity (a free operant relation), and in both, the activity itself was able to sustain a high rate of responding. Thus a tentative conclusion is that those tasks where students controlled the response rate and
where the task was capable of producing high rates of responding
are able to be significantly improved by the addition of the
sprint treatment, and in particular the peer sprint treatment.

Though Beth's data in Figure 15 was similar to the other
students it differs in several ways. First the levels of
responding for all correct responses are the lowest (with the
exception of the peer sprint, correct responses are below 25
rpm). The peer sprint appears to be the best treatment for Beth.
Though the mean incorrect rate was the highest of her incorrect
responses (and the highest of all students) at 14 rpm the peer
sprint produced a tripling of the rpm from baseline (8.7 to 28.7
rpm). Beth's data were also different from the other students
because correct responses for both baseline and the sprint
treatment crossover the data paths for the incorrect responses.
On Day 5 during baseline and Days 8 & 9 during the sprint
treatment, Beth had more errors than she had correct responses.
The sprint treatment for Beth was in these instances counter-
therapeutic. The variable most likely related to errors exceeding
successes was probably skill level. If Beth can’t perform the
skill she will produce errors. If the technique/ topography isn’t
remediated by increasing the rate of response (which was
doubtful), then it is reasonable that errors will increase. There
was throughout the volleyball study some remediation of errors
by Ms. East, the efficacy of which remains undetermined in this
study.
Fed Drill.

Data displayed in Figure 16 presents similar patterns for each student. Peer sprints was clearly the most effective treatment for the fed drill. The rates are low (below 25 rpm) relative to the wall drill, where the lowest rate obtained was 24 rpm. Incorrect response rates were low during baseline (see Table 14) at approximately 1.0 rpm for all students and like the wall drill both treatments produced a doubling (to 2 rpm) of the baseline incorrect response rate. Such low incorrect rates appear to be a product of the task and not the intervention. Interestingly the greatest gains relative to baseline were made in the fed drill. Jeff had a twofold increase in correct responding from baseline during the sprint treatment and a fivefold gain during peer sprints. Likewise Beth who scored only 2 correct responses during baseline doubled her correct responses during sprints and had a sixfold gain during the peer sprints. Gains in incorrect responses when compared to gains in correct responses were the lowest of all tasks and were lower in proportion to correct response gains.

1 V 1 Drill

Incorrect response rates were higher during the 1 V 1 drill than for the fed drill, but lower than those reported in Table 14 for the wall drill. Table 14 indicates that baseline levels for each student did not exceed a mean of 2 rpm and intervention
means did not exceed 5 rpm. Correct responses were most variable during sprint conditions (see Figure 17) for Jeff, Ian and Susan. Data paths showing ascending and descending trends for both treatments for all students suggesting that the data were not stable. Peer sprints were consistently higher than sprints (1 crossover for Ian on Day 7). Baseline rates of correct responding for Jeff, Ian and Susan also confound these results since the baseline gains for Jeff and Susan, in particular on Day 2, show marked improvement from Day 1. Induction (i.e., generalization) is the most likely explanation for this occurrence since on this day the treatment preceded the baseline tasks.

**Games Analysis**

Data on response patterns were once again similar across target students differing in magnitude, but not trend. The number of responses made by the team increased as team size increased. Jeff a high skilled male received more than the expected 33% contact with the ball when games were 3 X 3. He received less when the games were 5 X 5. Ian the low skilled male, met criterion during 3 X 3 if averaged over the 4 days. During 4 X 4, 5 X 5, and 6 X 6, games he exceeded criterion. Susan the high skilled female exceeded criteria on 5 X 5, and 6 X 6, games, but had a varied level of opportunity during 3 X 3 games. Beth the low skilled female failed to reach criterion on all except Day 1 where responses were extremely low. In general Beth was at 50%
or less of the criterion throughout the study without regard to the size of the team (though only 3 X 3, and 4 X 4 games were played. She was absent for the 6 X 6 games). The team membership changed daily, though frequently pairs of friends played together. Thus, the team data are not an artifact of a particular group of students, but are rather representative of the varying group membership in terms of ability and team size.

Class 2: Results

During this study all target students were absent from class at one time or another (John on Day 9, and Linda on Days 6 & 8, Tim & Carol on Day 3). Figure 19 displays the percentage of class time spent on lesson content for Class 2. Data for each category are relatively stable. The mean for general content was 44% (range 28%-54%). The knowledge category displays the most variability with a range from 13% on Day 11, from 52% on Day 1 (mean 27%). Data ranged between 20% and 43% for activity with a mean of 29%.

Three tasks (1 informing, 1 extending and 1 refining all serve related) were used by Ms. West. Like Ms. East, these were the only additional tasks used in the unit other than games and the tasks used in the treatments. Sixty-two percent of tasks were fully explicit, while the remainder (38%) were partially explicit. Tasks involving the treatments were usually fully explicit, most tasks involving games were partially explicit.
Figure 20 presents a direct comparison between the treatments (excluding baseline) measuring correct responses per minute for target students across tasks. The means and ranges of correct responses across treatments and tasks are presented in Table 15.

Table 15

Means and Ranges of Correct Responses for each Task Across Treatments for each Target Student in Class 2.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>John</th>
<th>Tim</th>
<th>Carol</th>
<th>Linda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wall Drill</td>
<td>50</td>
<td>56</td>
<td>38</td>
<td>43</td>
</tr>
<tr>
<td>Sprint</td>
<td>39-64</td>
<td>38-70</td>
<td>29-44</td>
<td>41-47</td>
</tr>
<tr>
<td>Wall Drill</td>
<td>51</td>
<td>56</td>
<td>42</td>
<td>45</td>
</tr>
<tr>
<td>Peer Sprint</td>
<td>44-64</td>
<td>48-62</td>
<td>38-42</td>
<td>40-50</td>
</tr>
<tr>
<td>Fed Drill</td>
<td>10</td>
<td>16</td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>Sprint</td>
<td>14-17</td>
<td>*</td>
<td>16-22</td>
<td>9-15</td>
</tr>
<tr>
<td>Fed Drill: Peer</td>
<td>16</td>
<td>18</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Sprint</td>
<td>9-20</td>
<td>16-20</td>
<td>16-18</td>
<td>11-20</td>
</tr>
<tr>
<td>1 V 1 Drill</td>
<td>9</td>
<td>14</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Sprint</td>
<td>*</td>
<td>13-15</td>
<td>8-9</td>
<td>8-9</td>
</tr>
<tr>
<td>1 V 1 Drill</td>
<td>18</td>
<td>19</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Peer Sprint</td>
<td>16-20</td>
<td>17-20</td>
<td>10-15</td>
<td>10-14</td>
</tr>
</tbody>
</table>

* signifies a range of zero
The means of incorrect responses for each task across baseline and treatment conditions are presented in Table 16.

Table 16  
**Means of Incorrect Responses for each Task Across Baseline and Treatment Conditions for each Target Student in Class 2.**

<table>
<thead>
<tr>
<th>Task</th>
<th>Conditions</th>
<th>John</th>
<th>Tim</th>
<th>Carol</th>
<th>Linda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wall Drill</td>
<td>Baseline</td>
<td>5.6</td>
<td>4.3</td>
<td>6.7</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>8.0</td>
<td>5.3</td>
<td>6.5</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>5.3</td>
<td>2.8</td>
<td>9.3</td>
<td>8.0</td>
</tr>
<tr>
<td>Fed Drill</td>
<td>Baseline</td>
<td>1.5</td>
<td>1.0</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>1.0</td>
<td>1.5</td>
<td>1.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>1.5</td>
<td>1.0</td>
<td>1.5</td>
<td>1.7</td>
</tr>
<tr>
<td>1 V 1 Drill</td>
<td>Baseline</td>
<td>5.3</td>
<td>2.7</td>
<td>5.3</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>5.0</td>
<td>1.5</td>
<td>3.5</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>3.8</td>
<td>3.8</td>
<td>2.5</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Figures 21-23 present within-task comparisons of baseline and treatments using correct and incorrect responses for the wall drill (Figure 21), the fed drill (Figure 22) and the 1 V 1 drill (Figure 23). In all Figures (21-23) treatments and
Figure 19. Percentage of class time spent on lesson content for Class 2 during the volleyball study.
Figure 20. Correct responses per minute for target students in Class 2 for Sprints and Peer Sprints across instructional tasks.
Figure 21. Correct and incorrect responses for the wall drill for target students in Class 2 during the volleyball study.
Figure 22. Correct and incorrect responses for the fed drill for target students in Class 2 during the volleyball study.
Figure 23. Correct and incorrect responses for the 1 V 1 drill for target students in Class 2 during the volleyball study.
Figure 24. Percentage of team responses made by target students in Class 2 (left y axis - histogram) and number of team responses (right y axis - freq. polygon) for students during the volleyball study. Proportion of team responses expected for the team size is indicated by horizontal dashed lines.
baseline measures of the tasks began on Day 4 (exception was
the fed drill baseline which began on Day 1. See Figure 22). Data
relative to the effects of team size on the proportion of
response made by a target student are displayed in Figure 24
(histogram display). The proportion of team responses expected
relative to team size were indicated by the dashed line and
identified by the arrow/percentage. The total number of
responses made by the team was represented by the frequency
polygon (solid line).

Class 2: Discussion

Time Use and Tasks

Unlike Study 1, the arrangement of time for Class 2 was
more stable throughout the Study 2. While the mean for general
content remains much the same for each study at 44%, the means
for knowledge and activity have exchanged 10% of the lesson
time. In Study 1 the mean for knowledge was 17%, in Study 2 it
increased to 27%. In Study 1, the mean for activity was 39%, in
Study 2 it decreased 29%. This rearrangement of time might be
explained a number of ways. Ms. West had previously played and
coached volleyball and she took a more active role in teaching
during the team taught lesson she shared with Ms. East. It was
the investigators observation that during the team taught
lessons observed (approximately half of those taught during the
investigation), the majority of the instruction regarding the
serve, the set and the bump was taught by Ms. West. Ms. West appeared to be more in control of Class 2 during volleyball than she was during the basketball unit. The three additional tasks reported in the results were serving tasks taught by Ms. West to all her classes (team taught and Class 2). These tasks were always explicitly stated as were the treatment tasks. Game tasks were for the most part partially explicit without any criterion being stated. When a criterion was stated it referred to when to change sides or end the game such as "After 4 points are obtained change sides".

Direct Comparisons Between Treatments Across Tasks

Figure 20 presents the data for correct responses across tasks for each treatment. The results reproduce the findings reported for Class 1. Sprints and peer sprints when used with the wall drill produce the highest rates of responding. The fed drill was marginally better than the 1 V 1 drill for all students except Tim. For Tim, the fed drill and the 1 V 1 drill are indistinguishable from the other in either treatment. Findings distinguishing between the peer sprint and sprint treatments are less clear in this study. There was crossover of the data paths for two of the students (Carol and Linda) for each treatment. Tim's data indicates that for him sprints were more effective than peer sprints. The data for John, Carol, and Linda provide no evidence of superior rates of responding for either treatment.
Thus though the effects in terms of rates relative to each task and treatment are reproduced, the finding from Class 2 regarding the effectiveness of the peer sprints is unsupported.

Wall Drill

The most noticeable finding from Figure 21 was that for Tim sprints were consistently more effective than peer sprints. Tim's means for correct responses presented in Table 15 are equal at 56 rpm for each treatment. However, Figure 21 shows a data path for the sprint treatment (Tim) that was ascending and increasing the variance between itself and the peer sprints data as time progresses. It would be wrong to conclude on the basis of the means in Table 15 that the treatments are equally as effective for Tim. In this case they are not. This finding was not supported by Class 1, which suggested that peer sprints would be more effective, nor was it supported by the data obtained from the other three target students in Class 2.

Rates of incorrect responding for students are below 10 rpms, but vary across students and treatments (see Table 16). For Carol the mean baseline rpm for incorrect responding was 6.7, the mean during the sprint condition was 6.5, and the mean during peer sprints was 9.3. Similarly, the baseline rpm for John was 5.8, during sprints it rose to 8.0, and during peer sprints it descended to 5.3.
Why the mixed findings? That the treatments are effective is clearly supported by the data presented in both studies to date. It is most likely that the variances are a product of differential responding or to use a modified colloquialism “different tasks and treatments for different folks”.

Fed Drill

As was the case for Class 1 the lowest (almost error free) responses were made during the fed drill (see Table 16). With the exception of Linda (baseline 1.6 rpm, sprint 3.0 rpm) both conditions produced nearly identical incorrect response rates as did baseline. Since mean baseline rates were approximately 1.5 rpm this was considered the best demonstration of a reduction in incorrect responding for both classes. The peer sprint treatment produced the highest levels of correct responding for John, Tim, and Linda. For Carol, the best treatment was the sprint. Baseline rates for correct responding for John, Tim, and Carol rose sharply after Day 1. Indeed the rates of correct responding for John and Carol, both high skilled students approached the lower rates produced by the treatment conditions. Though gains in baseline performance for correct responding began prior to the implementation of the treatments, the greatest gain was made from Day 4 to Day 5. The first introduction of the sprint condition occurred on Day 4.
Two conclusions are possible. One conclusion is that the
gains are a product of practice effects. If true then practice
affects would affect treatments and baseline equally, and any
differences between treatments and baseline would still be a
function of the treatments. A second possible conclusion is that
there was some induction or generalization occurring. On
Day 5 the baseline preceded the treatment, leaving only the
previous days treatment as an explanatory variable for the
induction. Induction is possible, but unconfirmed. There was not
a similar change in the data for Linda. In either case, for the
goal of comparison between treatments, neither explanation is a
threat to internal validity. Though conclusions about gains
relative to baseline for John and Carol need to be tempered with
the unexplained gains in baseline responses.

1_V_1_Drill

Baseline probes for this task occurred on Day 4 (see Figure
23). Intervention began on Day 6. There were only 2 occasions
when the sprint condition was introduced to the 1 V 1 task. For
each student (except John only 1 data point, due to his absence)
the data are indicative (only 2 data points) of an ascending data
path. Comments about the efficacy of the peer sprint condition
relative to the sprint condition need to be qualified with
reference to the few data points collected on the sprint
condition. Based on the data displayed in Figure 23 the peer
sprint treatments were the more effective of the two treatments in elevating correct levels of responding. Both treatments produced low rates of incorrect responding similar to those reported in Class 1 for this drill.

Games Analysis

Unlike Class 1, Class 2 had only two arrangements of team size. Teams were initially 3 X 3 (up to Day 7) and then were grouped into regulation 6 X 6 groups (Day 8 onwards see Figure 24). Like Class 1, when the number of members of the team increased so too did the total amount of responses made by the team. (The data presentation was identical to that reported for Class 1). John a high skilled student met or was close to criterion during the 3 X 3 games, but met criterion only once during the days when 6 X 6 games were played. Tim met criterion in both team configurations. Carol was more successful in meeting criteria during 6 X 6 games, than during 3 X 3 games. Seventy-five percent of the time Linda did not meet criterion during 6 x 6, on Days 6 and 7 she barely reached a fifth of the criterion. Linda was more successful during 6 X 6 games were she met criterion 50% of the time. For Linda a low skilled student games provided little opportunity for her to receive a “equitable share” of the opportunities to respond.
Study 2: Conclusions

Conclusions relative to Study 2 will be discussed under three headings. Conclusions relative to treatments, tasks, and games correlated with gender and skill level of the target students and reported across classes.

Conclusions Regarding Treatments

1. In this study the treatments first investigated in Study 1 were replicated using the same classes in a volleyball setting. In general, the effects for the sprint and peer sprint treatments reproduced the findings reported in Study 1. Correct responding made gains at least double that obtained in baseline. Specific findings are:

   a. Peer sprints produce higher levels of correct responding than sprint during wall drills.

   b. Sprints and peer sprints produce similar effects during fed drills and 1 V 1 drills. Gains are not of the same magnitude as found in the wall drill. Neither treatment demonstrated superiority to the other during either the fed drills and 1 V 1 drills.

   c. Neither sprints nor peer sprints, increase error rates in greater proportion than the increases in correct responding in any task studied. There was some evidence that at times a decrease in errors (both proportional and total) occurred from baseline during treatments.
d. Gender was not a variable that discriminated between treatments or task.

e. Skill level was a variable that discriminated between treatments and tasks. In short, low skilled students generally produced lower levels of responding than did high skilled students during treatments and across tasks.

f. There was some evidence that suggests that different tasks are appropriate for different levels of skill or task mastery.

Conclusions Regarding Tasks

1. In baseline and during both treatments the error rates were lowest for the fed drill. Levels of responding during this drill did not exceed 25 rpm in either class. This finding and the nature of the activity suggests that the fed drill was not a task that could produce high rates of responses, but was a task that could produce low error rates and increase the accuracy of the response. For low skilled students such as Beth in Study 1, the fed drill produced the most improvements over baseline. During the peer sprint condition Beth had a sixfold increase in correct responding. This finding suggests that for low skilled students the fed drill might be most suitable in developing skillfulness (i.e., correct typography)

2. The wall drill produced the highest levels of responding and the most errors when compared to the fed drill and the 1 V 1
drill. There was a relationship between performing the fed drill well and performing well on the wall drill. Those students in either study who had the higher fed drill correct responses and lower fed drill errors were also the students who performed well on the wall drill. In short, for students who could perform the fed drill well, the wall drill built endurance.

3. The 1 V 1 drill produced the most game-like responses with one student’s responses dependent upon the other student’s responses who in turn was dependent upon the first student’s responses and so on. The 1 V 1 drill was not capable of producing high levels of responding due to the potential for errors. Not surprisingly the 1 V 1 drill did produce the second highest amount of errors.

Conclusions Regarding Games

1. Skill level may influence an equitable share of responses. If you are a high skilled player (male or female) you are likely to receive more than an equitable share of responses (or at least equitable) than if you are a low skilled female. Low skilled females in each study made fewer responses, during any of the configured games than did their male counterparts. Additionally, they were at least 50% below criterion, for 50% of the lessons. This finding suggests that the inequitable share of responses was a product of gender and low skill level. The low skilled females in each study were less skilled than their low
skilled male counterparts. Beth in Study 1 was considerably less skilled than Linda or her male colleagues.

Skill levels were chosen by the teacher. Low skilled was a broad classification that teachers used to differentiate between students in basketball. Teachers were not asked to judge skill levels in volleyball (or badminton). Thus the classification of low skilled or high skilled may not hold across content. For example in Study 2 the low skilled male (Tim) received a higher percentage of responses in any team configuration than did the high skilled male (John) suggesting that though he was low skilled in basketball, he was not low skilled in volleyball. Data obtained during the interventions support the conclusion that Beth was low skilled in volleyball, that Linda was more skilled, but still low skilled, and that male students (designated by the teachers as high and low skilled) were as skilled as the high skilled females.

Thus low skilled students (determined by the nature and their responses) in the volleyball study were female. Hence the conclusion is that the differences in responses between students in the games setting was a function of the degree of skill and not gender.

2. The number of responses made by teams increased as team size increased.
The Badminton Study

Framing the Study

The goal of study 3 was to systematically replicate the interventions used in the first two studies using badminton as a new setting. Additionally, two of the tasks used during the volleyball study were modified for use in badminton. The fed drill and 1 V 1 tasks were examined using correct and incorrect response measures for each treatment. These drills had not been used by the teachers in their past badminton lessons with other classes (and were not used with their other concurrent classes). Their use was at the request of the investigator, though their modification was in cooperation with the teachers. Both teachers had requested that the time to administer the treatments be as minimal as possible due to the short period of time allocated for the badminton unit which was only 9 days. Thus only one of the tasks (the fed drill) had baseline data collected.

Game analysis similar to that conducted during Study 2 was also conducted in Study 3. Though the team sizes remained the same throughout, the focus remained on level of responding relative to the team's total responding (i.e., students receiving an equitable share).
Class 1: Results

Figure 25 displays the percentage of class time spent on lesson content for Class 1. Data for each category with the exception for Days 1 and 2 are relatively stable. The mean for general content was 36% (range 25%-48%). The mean for the knowledge displays was 24% with a range from 8% to 63%. Data ranged between 13% and 58% for activity with a mean of 41%.

No additional tasks other than the interventions and the 2 X 2 games were used. All treatment tasks (i.e., fed and 1 V 1 drills) were fully explicit, while tasks involving games were fully explicit 52% of the time and partially explicit 48% for the remainder.

Figure 26 presents a direct comparison between the treatments (excluding baseline) measuring correct responses per minute for target students across tasks. The means and ranges of correct responses across treatments and tasks are presented in Table 17.

The means of incorrect responses for each task across baseline and treatment conditions are presented in Table 18. Figures 27 and 28 present within-task comparisons of baseline and treatments using correct and incorrect responses the fed drill (Figure 27) and treatments only the 1 V 1 drill (Figure 28). In all Figures (26 -28) measures of the treatments-tasks began on Day 2 (exception was the fed drill baseline which began on Day 1. See Figure 27).
Table 17

Means and Ranges of Correct Responses for each Task Across Treatments for each Target Student in Class 1.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Jeff</th>
<th>Ian</th>
<th>Susan</th>
<th>Beth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fed Drill:</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Sprint</td>
<td>7-13</td>
<td>8-14</td>
<td>11-17</td>
<td>4-11</td>
</tr>
<tr>
<td>Fed Drill: Peer</td>
<td>16</td>
<td>13</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Sprint</td>
<td>13-20</td>
<td>7-16</td>
<td>8-16</td>
<td>3-13</td>
</tr>
<tr>
<td>1 V 1 Drill:</td>
<td>16</td>
<td>16</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Sprint</td>
<td>11-18</td>
<td>11-21</td>
<td>9-21</td>
<td>8-16</td>
</tr>
<tr>
<td>1 V 1 Drill: Peer</td>
<td>21</td>
<td>19</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Sprint</td>
<td>16-24</td>
<td>13-21</td>
<td>13-24</td>
<td>4-12</td>
</tr>
</tbody>
</table>

Table 18

Means of Incorrect Responses for each Task Across Baseline and Treatment Conditions for each Target Student in Class 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>Conditions</th>
<th>Jeff</th>
<th>Ian</th>
<th>Susan</th>
<th>Beth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fed Drill</td>
<td>Baseline</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>1.5</td>
<td>2.8</td>
<td>2.0</td>
<td>4.0</td>
</tr>
<tr>
<td>1 V 1 Drill</td>
<td>Sprint</td>
<td>2.5</td>
<td>3.5</td>
<td>2.8</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>2.3</td>
<td>2.8</td>
<td>2.3</td>
<td>3.5</td>
</tr>
</tbody>
</table>
Class 1: Badminton

Figure 25. Percentage of class time spent on lesson content for Class 1 during the badminton study
Figure 26. Correct responses per minute for target students in Class 1 for Sprints and Peer Sprints across instructional tasks.
Figure 27. Correct and incorrect responses for the fed drill for the target students in Class 1 during the badminton study.
Figure 28. Correct and incorrect responses for the 1 V 1 drill for the target students in Class 1 during the badminton study.
Figure 29. Percentage of team responses made by target students in Class 1 (left y axis - histogram) and number of team responses (right y axis - freq. polygon) for students during the badminton study. Proportion of team responses expected for the team size is indicated by horizontal dashed lines.
Data relative to the effects of team size of the proportion of response made by a target student are displayed in Figure 29 (histogram display). The proportion of team responses expected relative to team size were indicated by the dashed line and identified by the arrow/percentage. The total number of responses made by the team were represented by the frequency polygon (solid line).

Class 1: Discussion

Time Use and Tasks

In badminton, unlike basketball and volleyball, the activity time in the lesson with the exception of the tasks used for the treatments was devoted to game play. In discussions, Ms. East had indicated that she did little if any instruction during this unit other than during the first few days. The remainder of the time was devoted exclusively to tournament play. This emphasis was reflected in the data. The knowledge category occupied 38% and 63% of the lesson time during Days 1 and 2 and then drops to below 21% of the lesson time for the remainder of the unit. This 21% was devoted to mostly to the implementation of the treatments. The mean for general content during badminton (36%) was the similar to the means of the other 2 studies (basketball and volleyball were each 39%). Since knowledge was reduced to below 21% of the lesson (after Day 2) and the mean for general content was 36% there was more time available for
activity. The data reflect that the time devoted for activity averaged 41% of the lesson time. When a mean was recalculated excluding the Day 1 and 2 (low activity days due to instruction on rules and safety) the percentage of time was approximately half (48%) of the total lesson time. This was a substantial difference from previous means obtained in basketball and volleyball where the mean for each was 35%. The absence of tasks other than those used for treatments represents another significant difference from the instructional arrangement found in basketball and volleyball. During basketball and volleyball, the focus was clearly on skill development. During basketball other than the choice option (for a 3 X 3 game) there was no opportunity for game play. During volleyball games were introduced, but skill development was still a primary focus. During badminton team membership was fixed, (it was arbitrary for volleyball), scores were kept and tournament play was the focus. Strategic play was not emphasized during badminton.

Direct Comparisons Between Treatments Across Tasks

The most striking feature of the data presented in Figure 26 was the extent of crossover between data paths. The data for both treatments for each task display variability. Conclusions about the efficacy of drills and treatments are hindered by frequent crossovers of the data paths, variability, changes in trends, and non replicated effects across subjects. The mean
correct response data presented in Table 17 while reflecting marginal differences between mean also describes variability in the range. Conclusions about differences between treatments should not be based on the means. For example were the data in Figure 15 scaled using a y axis from 0 to 100 the data paths would be indistinguishable from each other. There are no comparative conclusions warranted from these data, other than for badminton the treatments produce a variable rate of responding and that each treatment produced similar effects.

Fed Drill

As has been the case in Study 1 and 2, Beth's data differs from the data displayed for Jeff, Ian and Susan. The data for Jeff, Ian and Susan show low almost errorless learning for baseline and treatment conditions. In contrast, Beth's data for incorrect responses indicates initially more incorrect responses for both treatments than there were correct responses. This persisted for 4 days (Day 2 through 5) then reversed itself. There was evidence in Figure 27 that peer sprints appeared to be the better treatments for Jeff and Beth, and that sprints were the better treatment for Ian and Susan. This finding supports the earlier conclusion from Study 2 regarding the differential effects of treatments.
L_V_1_drill

Similar findings for the 1 V1 drill as those found for the fed drill will be reported here. There was no clear evidence from Figure 28 that one treatment was better than another. There were no baseline probes taken during the 1 V 1 drill (due to teacher requests for time to maximize game play). Thus findings relative to baseline are unable to be made. The incorrect responses for Ian, Susan and Beth are higher than similar rates on this drill in volleyball. For Beth the rate of incorrect response did not reduce as time passed. Findings comparing treatments support the conclusion of differential effects.

Game_Analysis

Data on response patterns for the target students is reported in Figure 29. In the volleyball study a conclusion was stated that “The number of responses made by team increased as team size increases.” In this study team size was constant (2 X 2). Overall team responding did not significantly increase other than the odd data point rising (as in Ian’s team Day 6) or falling (as in Jeff’s team Day 4). Trends for total team responses were ascending but not significantly different on Day 9 from Day 3.

Of 7 days of badminton games, Jeff a high skilled male, Ian the low skilled male, and Susan the high skilled female met or exceeded criterion on 5 of the 7 days. Beth the low skilled female met criterion on only 2 days, though she fairied
considerably better than she did during volleyball. Perhaps this more equitable sharing was a feature of the forced choice in responding in a 2 X 2 game. When a shuttlecock flies over the net it was going to land on one side of the court, or the other if it lands in (50% opportunity to respond for each student presumably). The 2 X 2 game might be a better arrangement in volleyball than the other team configurations (e.g., 3X3 or 6X6). The team membership did not change daily, but opposing teams did. The total amount of responding during badminton given the available time was quite low. This was due, in part to the number of teams and the number of badminton courts. At any given time at least a third of the class was changing courts, sitting out waiting for a free court or giving game scores to the teacher.

Class 2: Results

Figure 30 displays the percentage of class time spent on lesson content for Class 1. The mean for general content was 35% (range 28%-45%). The mean for the knowledge category was 24% with a range from 8% to 55%. While the mean for activity was 42% and a range from 8% to 65%.

No additional tasks other than the interventions and the 2 X 2 games were used. All treatment tasks (i.e., fed and 1 V 1 drills) were fully explicit while tasks involving games were
partially explicit 56% of the time and fully explicit for the remainder (44%).

Figure 31 presents a direct comparison between the treatments (excluding baseline) measuring correct responses per minute for target students across tasks. The means and ranges of correct responses across treatments and tasks are presented in Table 19. The means of incorrect responses for each task across baseline and treatment conditions are presented in Table 20. Figures 32 and 33 present within-task comparisons of baseline and treatments using correct and incorrect responses the fed drill (Figure 32) and treatments only the 1 V 1 drill (Figure 33). In all Figures (31 -33) treatments and baseline measures of the tasks began on Day 2 (exception was the fed drill baseline which began on Day 1. See Figure 32). Data relative to the effects of team size of the proportion of response made by a target student are displayed in Figure 34.
Table 19

Means and Ranges of Correct Responses for each Task Across Treatments for each Target Student in Class 2.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>John</th>
<th>Tim</th>
<th>Carol</th>
<th>Linda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fed Drill:</td>
<td>15</td>
<td>14</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Sprint</td>
<td>9-19</td>
<td>10-16</td>
<td>11-20</td>
<td>10-20</td>
</tr>
<tr>
<td>Fed Drill: Peer</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>Sprint</td>
<td>13-20</td>
<td>13-20</td>
<td>14-22</td>
<td>11-18</td>
</tr>
<tr>
<td>1 V 1 Drill:</td>
<td>21</td>
<td>22</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Sprint</td>
<td>16-27</td>
<td>17-28</td>
<td>16-21</td>
<td>16-22</td>
</tr>
<tr>
<td>1 V 1 Drill:</td>
<td>18</td>
<td>19</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Peer Sprint</td>
<td>10-24</td>
<td>9-26</td>
<td>11-26</td>
<td>12-25</td>
</tr>
</tbody>
</table>

Table 20

Means of Incorrect Responses for each Task Across Baseline and Treatment Conditions for each Target Student in Class 2.

<table>
<thead>
<tr>
<th>Task</th>
<th>Conditions</th>
<th>John</th>
<th>Tim</th>
<th>Carol</th>
<th>Linda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fed Drill</td>
<td>Baseline</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Sprint</td>
<td>1.8</td>
<td>1.3</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>2.0</td>
<td>1.5</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>1 V 1 Drill</td>
<td>Sprint</td>
<td>2.3</td>
<td>1.5</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Peer Sprint</td>
<td>2.5</td>
<td>2.3</td>
<td>4.5</td>
<td>2.8</td>
</tr>
</tbody>
</table>
Figure 30. Percentage of class time spent on lesson content for Class 2 during the badminton study.
Figure 31. Correct responses per minute for target students in Class 2 for Sprints and Peer Sprints across instructional tasks.
Figure 32. Correct and incorrect responses for the fed drill for the target students in Class 2 during the badminton study.
Figure 33. Correct and incorrect responses for the 1 V 1 drill for the target students in Class 2 during the badminton study.
Figure 34. Percentage of team responses made by target students in Class 2 (left y axis - histogram) and number of team responses (right y axis - freq. polygon) for students during the badminton study. Proportion of team responses expected for the team size is indicated by horizontal dashed lines.
Class 2: Discussion

Time Use and Tasks

Ms. West followed an almost identical instructional arrangement to Ms. East. Time use data and tasks are highly similar reflecting the influence of the team taught lessons as models for Ms. West. Like Class 1, the entire lesson with the exception of the treatments-tasks was devoted to game play. Like Class 1, this emphasis was reflected in the data. The knowledge category occupied 48% to 55% of the lesson time during Days 1 and 2 and then drops to below 23% after Day 4 for the remainder of the unit. Knowledge during Day 4 to 9 was devoted mostly to the implementation of the treatments. The mean for general content during badminton (36%) was the lowest for the three studies (basketball and volleyball were each 44%). Perhaps there was less to organize during badminton in terms of drills and tasks (e.g., time for instruction and demonstrations). Since knowledge was reduced to below 23% of the lesson after Day 4 and the mean for general content was 35% there was more time available for activity. The data reflect that the time devoted for activity averaged 41% of the lesson time. When a mean was recalculated excluding the Day 1 and 2 (low activity days due to instruction on rules and safety) the percentage of time was exceeds half of the total lesson time (55%). As was
the case with Class 1 this was a substantial difference from the previous means obtained for volleyball and badminton.

Direct Comparisons Between Treatments Across Tasks

Similar to the Class 1 data, the most striking feature of the data presented in Figure 31 was the extent of crossover between data paths. The data for both treatments display variability across tasks. Conclusions about the efficacy of the treatments are confounded by frequent crossovers of the data paths, variability, changes in trends, and non replicated effects across subjects. Like Class 1 there are no comparative conclusions warranted from these data. However the replication of negligible difference between treatments, and the similarity of student responses across classes, suggests that the effects may have been an artifact of the instruction used by both teachers.

Fed Drill

The data for John, Tim, Carol and Linda reported in Figure 32 show low, almost errorless learning for baseline and treatment conditions, though only 1 baseline probe was conducted. Compared to the baseline probe both treatments produce gains at least twofold. These findings reproduce the Class 1 results and support the earlier conclusions from Study 2 regarding the differential effects of treatments. The pattern of the data path (change in
trends) was similar for each treatment across students. Initially peer sprints are more effective than sprints, then the sprint treatment (usually showing a ascending data path) produces more responding than the peer sprint treatment.

I V 1 drill

Similar findings to those found in Class 1 and in Study 2 for the I V 1 drill as those for the fed drill will be reported here. There was no clear evidence from Figure 33 that one treatment was better than another. Patterns are similar for John and Tim, (sprints initially more effective then later peer sprints producing higher levels of responding. There were no baseline probes taken during the I V 1 drill (due to teacher requests for time to maximize game play). Thus findings relative to baseline are unable to be made. The incorrect responses for all students reflect similar levels obtained for Class 1. Findings comparing treatments are consistent across students differing only in magnitude but not effect. These findings though do not distinguish well between of the efficacy of the sprint treatment and the peer sprint.

Game Analysis

Data on response patterns for the target students is reported in Figure 34. Team response rates increased as time passed in this study. Thus the conclusion arrived at in the
volleyball study that “The number of responses made by team increased as team size increases” may be a function of more practice and not changing team sizes. Data paths for John, Tim, and Carol rose throughout the study. This finding was similar, and different to the effects reported in Class 1. In Study 2, changes in team size were correlated with increases in total responding. However increases in total responding in this study (Study 3) occurred over time suggesting the influence of practice effects, not team size (team size was constant during badminton).

Over 7 days of badminton games John and Carol both high skilled student achieved criterion in six lessons. John exceeded the criterion by at least 10% on 3 of the 7 days. Tim the low skilled male met criterion twice while Linda the low skilled female met criterion once though she was within 5% of criterion every other day. The team membership did not change daily, but opposing teams did. The total amount of responding during badminton was quite low. In part this was due to the number of teams and the number of badminton courts. Like Class 1, at any given time at least a third of the class was changing courts, sitting out waiting for a free court or giving game scores to the teacher.
Study 3: Conclusions

Study 3 sought to experimentally replicate the treatments used during Studies 1 and 2, and to replicate these treatments across tasks used during Study 2. Additionally, an investigation into equitable responses in games, begun in the volleyball study, was continued in the badminton setting. The conclusions relative to these objectives are:

1. The interventions were successfully replicated in the badminton settings however,
   a. Differences between interventions remain unclear. During the fed drill both treatments produced effects similar to those reported in Studies 1 and 2 namely twofold increases in levels of correct responding and low rates of incorrect responding. Neither intervention was superior to the other. This finding reported in Class 1, was replicated in Class 2.
   b. During the I V 1 drill no baseline measures were obtained and thus conclusions relative to baseline cannot be made. The interventions produced similar rates of responding to the fed drill. Similarly, there was no evidence to support a conclusion that one treatment was better at elevating rates of correct responding than the other.

2. When students were playing 2 X 2 games of badminton equity of responses was relatively equal. Skill level did it differentiate between the share of responses but only marginally so. Thus given conditions where a response should stand a 50%
chance of being made by one student or another, the conditions were more equitable in badminton, than in volleyball.

3. In volleyball the number of responses made by the team increased as the team size increased. However the team size increased as time passed (from Day 1 to Day 11). In badminton the team sizes were held constant and team responses still increased. Thus on the basis of this evidence a conclusion that changes over time in the number of responses made by a team might be as likely due to practice effects (i.e., teams became more efficient, successful and produced more opportunities due to practice) as they are might be due to changes in team size.

Comments on Study 3 Findings

The findings obtained from Study 3 offer support for some conclusions from Studies 1 and 2. The findings from Study 3 also differ from those obtained in Study 2. Since the interobserver agreement data on the dependent variables and on the procedural fidelity are reliable and assumed accurate, some explanation is warranted. The replication of negligible difference between treatments, and the similarity of student responses across classes, suggests that the effects may have been an artifact of the instruction used by both teachers in Study 3 that they did not use in the previous studies. There was evidence to support this proposition. The teachers in Study 3
differed in their arrangement of the instructional conditions for Studies 1 and 2 in the following ways:

1. The general picture was that badminton produced the highest percentage (approximately 45%) of lesson time devoted to activity, but the least amount of time devoted to instruction when compared to basketball and volleyball. Game play (in either class) was not interrupted for remedial instruction.

2. The focus was on game play from Day 1 in Study 3 which was in contrast to the emphasis on skill development observed during volleyball, and in particular in the basketball study. Both teachers who initially requested to reduce the amount of time (and therefore extent of the intervention) devoted to the treatments in favor of maximizing game play. Treatments though faithfully implemented, were seen by teachers as something to get out of the way so that game play could begin. An escape contingency was frequently articulated by both teachers during instruction (e.g., “do these drills first before playing games”). Since the only contingency relating the treatments to the game play produced an escape from the treatments it was possible that students were also more concerned with game play than the drills. If true these explanations might account for the lack of differential findings between treatments.

3. The drills intervened on in studies 1 and 2 were drills that the teachers commonly used in their classes and used in the
other classes they taught that were not a part of the investigation. During basketball and volleyball teachers used the tasks and some of the interventions in other classes. Neither teacher was observed doing this in the badminton units for these other classes.

The instructional conditions during basketball and the instructional conditions during badminton were diametrically opposed in terms of philosophy (skill development versus game play) and tasks. Given this difference a potential conclusion would be to consider the data in Study 3 as flawed. Such a conclusion would reflect on the internal validity of the study. There is no evidence to support this conclusion (see measures of procedural fidelity and interobserver agreement on the dependent variable reported earlier). Instead the findings indicate a problem with external validity. The conclusion may be stated as "Perhaps the treatments work best in conditions where there is a philosophy of skill development and work least effectively in conditions where game play is emphasized at the expense of skill development."

A additional explanation for the lack of difference between treatments might be multiple treatment interference. This was the third study that required students to interact with the treatments. Generalization of effects is a recognized phenomenon with alternating treatments designs (Ulman and Sulzer-Azaroff, 1975). It is possible, that the data in this study
is indistinguishable because of multiple treatment interference. Though possible (i.e., generalization) it is in the opinion of the investigator more likely a function of the instructional conditions.

**Generalization**

Stimulus generality is defined by Cooper, Heron and Heward (1987) as the extent to which the target behavior occurs in environments other than those in which the behavior was trained. In this investigation the teachers were the subjects, and the behavior was the intervention (more accurately the teaching behaviors used in the intervention). Generalization to other settings was measured by informal observations by the principal investigator and two assistants during the break between Class 1 (period 1) and Class 2 (period 4) each day. During this time the investigators waited in a side room from where they were able to observe the classes occurring in periods 2 and 3 team taught by Ms. East and Ms. West unobtrusively. The investigators' frequently observed that during Study 1 and 2 the drills and the sprint conditions were used as a regular part of the instruction. Specifically during Study 1 the investigators' observed the use of the sprint treatment by the teachers in both period 2 and period 3 lessons for six days. During Study 2 investigators' observed the use of the sprint treatment by the teachers 6 of the 8 days in which the treatments were used in Class 1 and Class 2.
No examples of the hustles treatment or the peer sprints were observed. When asked why the peer sprints were not used since the peer sprint was considered by the teachers to be the more effective of the two treatments, Ms. East responded that they didn't have spare coding cards, and that handing out the cards and pens took time. None of the drills used in badminton for the treatment were observed being used in the team taught lessons. It is appropriate to conclude that the sprint treatments did generalize of other settings initially, and that the hustles and peer sprints did not.

Social Validity

The generalization of the treatments provides some measure of the acceptability and hence social validity of the treatments as seen by the teachers. As explained in Chapter 3: consumer satisfaction and the acceptance of the procedures may be obtained if teachers were to use the interventions in other classes not under experimental control, or if teachers behaved in a manner consistent with their self reports in interviews, and if students “chose” from a menu of treatments, the treatment they preferred to operate in their classes.

Using the sprint treatment in other classes strengthens the case for the utility of that intervention. An intervention has utility if it is both easy to implement and practical for teachers
to do so. Hustles were neither easy, nor practical for teachers to implement, and peer sprints though practical, required a greater level of preparation (pens and cards) than sprints which produced similar results for less effort.

Social validity of the interventions from the perspective of the teachers was also evaluated through informal interviews. During debriefing at the end of each study, both Ms. East and Ms. West reported on the costs and benefits of the treatments. For example, while discussing the hustles treatment Ms. East commented “It is not something I could do every day or even for short periods in each lesson.” Ms. West agreed and both teachers suggested that they would prefer not to have to use it during the volleyball study. Ms. East and Ms. West both saw the sprint and peer sprint as beneficial. Ms. East commented on the peer sprint “I like this progression [referring to the treatment effects] more” and then discussed the importance of the peer sprint in teaching her students to discriminate between knowing and performing the technique of the skill.

After 2 days when she had not used the intervention (time used for testing) with the team taught classes Ms. West remarked “I miss it” referring to the higher rates of responding produce by sprints and peer sprints (i.e., “I miss the intensity”).

Students as end-line consumers of the treatments were also given an opportunity to assess the acceptability of the treatments (albeit a forced choice). In Study 1 students in each
class were asked to choose from among the three treatments, hustles, sprints and peer sprints which they would prefer. As discussed in Study 1 the group choice in both classes was for the sprint condition. Though there was some support for the peer sprint there was no support for the hustle. Teachers and students appear to agree on the acceptability of the sprint treatments.
CHAPTER V
SUMMARY, CONCLUSIONS, IMPLICATIONS
AND RECOMMENDATIONS

We found instructional settings that were at best, distinctly causal, with substantive learning gains modest. In the classes we found clear winners and losers, most frequently definable by gender and skill level (Siedentop & O'Sullivan, 1992, p. 285)

"I just don't believe you can push the average Joe" (Mrs. East 1 year before this investigation during an interview)

Summary of the Investigation

The above quotations suggest instructional conditions that are very different from (a) Rosenshine’s and Stevens’ (1986) description of a high paced elementary school math episode “It is exciting to watch the class or group move at a rapid pace and to watch all the students giving the correct response rapidly and confidently” (p. 380), or (b) from Tharp’s and Galimore’s (1976) use of descriptors such as intense, concentrated, continuous, and repetitious to describe coach and player
behavior during basketball practice sessions, or (c) Lindsley’s
description of precision teaching at work during a one minute
timing episode in a reading lesson,

The students are busy at their desks, in teams of two,
timing each other’s practice, jumping up to take a chart
down from the wall, or to post new data. The students are
noisy, shouting correct answers as fast as they can at 200
words per minute (p. 25).

The authors of these descriptions would recognize the
effects and possibly even the conditions that produced them if
asked to observe the settings for each of the other descriptions.
Tharp and Galimore (1976) saw the relation between the
instructional conditions they observed John Wooden use to
intensify his practices, and classroom situations. They noted
“If such concentrated effort ever finds its way into the ordinary
classroom, the results might be rewarding” (Tharp & Galimore
1976, p. 78). What is common to each of the previous
descriptions is a high rate of not just responding, but correct
responding. In short, getting it right frequently.

One of the tasks of Chapter 2 was to find evidence in favor
of rate of response as an alterable variable. The nature and the
rates of response where found to vary within and across
classes, content and teachers in both regular education and in
physical education. In physical education the review revealed
that while response rates have been primary or secondary
sources of data for studies, neither qualitative nor quantitative
dimensions have been directly intervened upon. There was however considerable descriptive-analytic as well as anecdotal evidence from these studies that suggested that the qualitative and the quantitative dimensions of responding in physical education can, and should be intervened upon with a view to improving the achievement of students in physical education.

This investigation was experimental because as Siedentop (1982) noted “If you want to really understand something, try to change it.” The purpose of this investigation was to modify selected instructional conditions that have been used in classrooms, which have been associated with producing high rates of correct responding. Three conditions were used in this investigation:

1. Hustle which was a instructional condition where high rates of teacher prompts and hustles were designed to increase rates of student responding.

2. Sprints made use of timed practice (sprints) where correct performances were counted, and where the focus on the instructional conditions was to produce as high a rate of responding as was possible for a particular context. In this condition student’s self reported to the teacher regarding the number of correct responses. The teacher’s instructional focus on the performing student in order to confirm the self reporting and to motivate students during the sprint.

3. Peer sprints were an identical condition to the sprint condition, except that peer sprints involved one member of the
pair keeping a record of the performance of the other member. In the peer sprint, one student discriminated and recorded the correct responses of his/her partner. The student observing reported to the teacher on their partner's success. The teacher's instruction focused on the observing/recording student in order to ensure that they were on task. The particular design arrangement involved collecting baseline data on students responding during teacher selected (and commonly) used tasks. These tasks were experimentally analyzed across treatments using an alternating treatments design. The dependent variables used in this investigation to judge the efficacy of the treatments were correct or incorrect (error) rates for the typography of the response and total response rates.

Originally the intention was to include response products as dependent variables (indeed the experimental questions remain phrased that way), however the video camera had to be zoomed close enough to detect the topography of the response, and keep all students in view of the lens, but still capture the outcome of a response. This proved impossible and rather than capturing an incomplete record, only measures on topography were used since these were the focus of the teachers' contingencies and instruction.

The treatments were first examined in a basketball setting then replicated in volleyball and badminton. In each setting a second class was used to provide a systematic replication of the experimental effects obtained for Class 1.
The findings from these three studies were presented and discussed in Chapter 4. This chapter will present the conclusions relative to the experimental questions and present recommendations for future research.

Conclusions

In this section conclusions relative to the experimental questions and conclusions related to the games investigation will be presented. Two broad umbrella questions framed the investigation: “What are the effects of modified classroom instructional treatments in physical education?” and “How does an increased rate of responding affect the behavior of students and teachers in physical education lessons?” Each of these umbrella questions is answered by their more specific sub-questions.

Answering the Experimental Questions

1. What are the Effects of Modified Classroom Instructional Treatments in Physical Education?

1.1 Do the treatments improve the response rate, success rate, and correct topography rate of the target students relative to “standard” practice? Yes, some more than others, dependent upon the type of task. The major finding from this investigation may be stated as: All treatments consistently produced at least a doubling of baseline rates for total response rates and for
correct responding in all situations with the exception of the hustles treatments when applied to passing. Proportional gains were reported as high as five and sixfold increases over baseline for some students. Increasing the response rate and the correct responding rates did not appear difficult for teachers nor aversive for students, though there were times the appropriateness of the treatments were questioned (hustle in basketball and sprints and peer sprints in badminton). The treatments were robust in that changes occurred immediately during intervention, and the changes were of a significant magnitude (at least twofold in size). Ms. East’s comment that “I just don’t believe you can push the average Joe” appears unsubstantiated for the tasks and settings investigated in these studies. The classes taught by Ms. East and Ms. West during the interventions did not fit the casual description used by Siedentop and O’Sullivan (1992) who reported “We found instructional settings that were at best, distinctly causal, with substantive learning gains modest” (p. 285). Rosenshine and Stevens, Tharp and Galimore, Lindsley or indeed any coach of a high achieving and successful sport team would have recognized the intense, task oriented, success oriented and achievement directed instructional conditions produced by the treatments.

Griffin (1991) reported in her conclusions regarding player responses in an interscholastic varsity volleyball team that “practice focused on mastery, utilizing repetitive practice tasks . . . . Another quality was the practice intensity reflected
by the high amount of motor engaged OTR's with high topographical appropriateness" (p. 123). Such a description could be equally applied to the treatment phases of the lessons observed in this investigation, though the description would not be relevant for the other phases of the lesson (game play and baseline tasks). Thus increases in the rate of responding during treatments producing a "practice [in this case lesson] intensity" followed by a lower intensity during other tasks in the lesson, suggest a role to be filled, and a way to view the treatments used in this study. These treatments might be best viewed as tactics used by teachers to boost intensity (rate of responding) in the same way that Tharp and Galimore (1976) described Wooden's efforts to intensify his practices. However they should not be viewed as strategies, (i.e., entire lesson consisting of 1 minute timings or hustles) but rather viewed as probes for practice and assessment, that help to inform the students and teachers regarding levels of achievement.

1.2 Are any of the treatments superior to the others relative to response rate, success rate, and correct topography rate? In Study 1 (basketball) there was some evidence (See Class 2 data) that suggests that peer sprints are more effective than sprints during dribbling tasks in producing more responses and more correct responses. The hustles treatment during dribbling tasks produced at times high, but variable responding
for both classes and the conclusion from Study 1 was that it was not consistently better than the other two treatments.

In Study 2, (volleyball) findings relative to the superiority of one treatment over the other differed by task (in this study only sprints and peer sprints were used). During the wall drill the peer sprints were clearly more effective in producing high rates of correct responding than the sprint condition. The difference was indeterminable between either treatment during the fed and the 1 V 1 drills.

In Study 3, the finding that the indeterminable difference between either treatment during the fed and the 1 V 1 drills was replicated. There were no distinguishing features of the data that allowed the effects of the treatments to be discriminated.

1.3 In a forced choice between treatments, which treatment is selected by the students? In Study 1 students in each class were asked to choose from among the three treatments, hustles, sprints and peer sprints, which they would prefer to use on the last day of class. The sprint condition was the preferred treatment in each class. Some students in each class did opt for the peer sprints, while none chose in favor of the hustles.
1.4. which (if any) of the treatments are used by the teachers in their other classes? The sprint treatment did generalize to other settings, however the hustle treatment and peer sprints did not. During the basketball and volleyball studies the sprint treatment was used by both teachers in their other classes. Neither the hustles nor the peer sprints were observed outside of Class 1 and Class 2. None of the treatments were used in other classes during the badminton study. Interviews conducted with the teachers suggest that the hustles treatment was not liked by the teachers for at least two reasons (a) it was physically demanding and (b) it focused their attention on a small component of instruction at the potential expense of others. Peer sprints were well liked but involved more organization (producing and handing out cards and pens). For less effort sprints delivered similar effects to the peer sprints. Interestingly both teachers and students appear to agree on the acceptability of the sprint treatments.

2. How does an Increased Rate of Responding Affect the Behavior of Students and Teachers in Physical Education Lessons?

2.1.1 Does an increased rate of responding affect high and low skilled students differently? A common finding from the research reviewed in Chapter 2 has been that high skilled students are more successful more often than low skilled students (Buck, Harrison, & Bryce, 1991; Graham, 1986; Lund,
In this investigation that finding has been supported. Skill level was a variable that discriminated between the magnitude of responding. That is low skilled students generally produced lower levels of responding than did high skilled students. Skill level was not, however, a variable that prevented significant gains. For example during Study 2, Beth a low skilled student in Class 2 scored only 2 correct responses during baseline, doubled her correct responses during sprints, and had a sixfold gain during the peer sprints.

Graham (1986) reported that low skilled students not only had lower total responses, but that those responses were more often incorrect. In general, that was not the case in this investigation. There were times when Beth’s incorrect responses exceeded her correct responses but this relationship was reversed 3 days into the interventions.

The previously reported finding that low skilled students generally produced lower levels of responding than did high skilled students is confounded by the definition of a low skilled student. “Low skilled” and “high skilled” were broad classifications that teachers used to differentiate between the skill levels of students in basketball. Teachers were not asked to judge skill levels in volleyball or badminton. Thus, the classification of low skilled or high skilled might not necessarily hold across content. In reality, it did not. For example, in Study 2 on the basis of the data for Tim the low skilled male (in basketball) one must conclude that he was
equally as skilled as the high skilled male in the group. Indeed Tim received a higher percentage of responses in any team configuration than did the high skilled male (John). Low skilled students in this investigation were ultimately defined by the data. Beth and Linda both identified as low skilled students were in fact the lowest skilled of the target students based on the data obtained for the three studies. The issue here is one of accuracy. Researchers defining low and high skilled students may need to pre-test and make some assessment of the degree to which a student is “low skilled.”

2.1.2. Does an increased rate of responding affect male and female students differently? Yes and no. The high rates of responding produced by the treatments affected males and females similarly in terms of trends (they all improved), but differently in terms of magnitude. Though treatments typically produced between two and sixfold gains in responding when compared to baseline low skilled female students produced low levels of responding when compared to their male counterparts even during treatments. This suggests that differences in the level of responding across students was a function of skill level and gender. This view is consistent with the observations of Siedentop and O'Sullivan (1992).
2.2 Does the explicitness of tasks stated by the teacher change as the rates of responding changes? The explicitness of the task statements were not affected by the changes in rates of responding. Most if not all task statements for the treatments were explicitly stated. Typically, skill instruction tasks were explicitly stated and games related tasks were partial explicit. These relations were not a product of the treatments, nor the rate of responding, but rather the way these teachers taught. The finding that both teachers used fully explicit task statements most often during instruction is consistent with findings reported by Jones (1989) and Lund (1990).

2.3 In what ways does an increased rate of responding affect how students and teachers spend their time in physical education lessons? There was no evidence from the time use data that changes in rate of responding affected the distribution of time between general content, knowledge or activity in any manner. In studies 1 and 2, Ms. East in particular, and Ms. West less so spent considerable time during instruction ensuring that students "knew" the task and could discriminate it. This appeared to be a normal part of the instructional repertoire of both teachers and was unrelated to the treatments.
Games

Interest regarding games was occasioned by the opportunity to collect data on responding during games. The strategy pursued examined the responses by a target student as a member of a team. These responses were compared relative to the size of the team. The basic question was “Do students receive an equitable share of responses during game play?” The question was investigated first during volleyball where the issue of an equitable share of responses was investigated as a function of 3 X3, 4 X 4, 5 X 5, and 6 X6 games in Class 1, and 3 X3, and 6 X6 games in Class 2, and in Badminton using the 2 X 2 game format.

Data obtained from Study 1 indicated that skill level may influence an equitable share of responses and that the number of responses made by teams increased as team size increased. Data obtained from Study 3 indicated that when students played a 2 X 2 game of badminton equity of responses was relatively equal. Skill level did differentiate between the share of responses, but only marginally so. Thus given conditions where a response should stand a 50% chance of being made by one student or another, the conditions were more equitable in badminton, than in volleyball. The conclusion reached suggests that the smaller the team size is the more equitable the responses. The most equitable condition was found during 2 X 2 badminton games where the opportunity to make a response was highest 50%.
In volleyball the team size increased as time passed (from Day 1 to Day 11). Responses also increased as time past leading to the tentative conclusion that the number of responses made by the team increases as team size increases. In badminton, the team sizes were constant, yet team responses increased. Thus on the basis of this evidence a conclusion that changes over time in the number of responses made by a team might be as likely due to practice effects as they might be due to changes in team size.

Implications

In 1984 Bloom described what he called the “2 sigma problem” namely, “Can researchers and teachers devise teaching-learning conditions that will enable the majority of students under group instruction to attain levels of achievement that can at present be reached only under good tutoring conditions” (Bloom, p. 4). This investigation examined three conditions designed to increase rate of response during group instruction. Rate of response was proposed as a functional (and alterable) variable directly influencing student achievement.

Rate of response proved indeed to be alterable, specifically rates of correct technical responding increased, and rates of incorrect technical responding were maintained at low levels, or reduced. The intervention produced significant and consistent gains across subjects, settings and tasks.
The conclusions of this investigation previously summarized in this chapter are more powerful than they might otherwise have been had only one study been performed. The findings and conclusions are the product of 3 studies replicated in different settings (basketball, volleyball and badminton) with the same subjects in two classes.

High school settings in physical education need not be places where students are unsuccessful. The students in this study male and female, low and high skilled, were able to substantially improve their performance on the tasks selected by their teachers. They did so with little if any complaint, or counter control. Granted, there is no evidence to suggest that the tasks directly improved game playing, but that is a question of curricula selection, and though important was not a focus of this study. The teachers though they expressed concern regarding the appropriateness of the hustle drill, and the use of sprints and peer sprints during badminton, were nonetheless supportive and enthusiastic about the treatments and their effects during basketball and volleyball. Notwithstanding the comments regarding badminton and the hustle treatment, changing the rate of response (and correct rates of responding) was not a difficult thing to do. It seems quite reasonable to suggest that the physical education high school environment need not be “distinctly causal, with substantive learning gains modest” (Siedentop & O'Sullivan, 1992, p. 285), indeed the reverse is true in this investigation. The environment was
intense and learning gains were far from modest. Altering rate of response, in particular correct rates of responding is a effective way to change the learning environment described by Siedentop and O'Sullivan (1992) and in the process move instruction in physical education closer to the goal of enabling the majority of students under group instruction to attain mastery of the subject matter.

The treatments/conditions as previously suggested in this chapter are best conceived of as tactics for increasing rates of response and rates of correct responding. The treatments should not be viewed as strategies, in the way that direct instruction (Rosenshine, 1979) is a strategy for academic success, though there are possibly two general rules that can be gleaned from this investigation: (a) produce a lot of correct responding and (b) keep errors low to nonexistent. These rules might be used as criteria to judge the efficacy instructional conditions used by teachers in teaching physical education. Instructional conditions using these two rules would not require predictor variables, but direct measurement of learning. As such measures of correct and incorrect responding and provide an direct and ecologically valid measure of student performance, which is itself a measure of student achievement.

Among the many findings derived from this investigation is the finding that the rate of responding varied across treatments, tasks, and students, and this was replicated across teachers and settings. This finding is a familiar one in the

The most common finding from these studies is that responding differs most frequently on the basis of skill level. The finding in this investigation that tasks produce variable responding suggests, as much of the literature has recognized, that different tasks might be more appropriate for different students and that the mediating variable is skill level. One way of categorizing tasks might be according to the type of responding they frequently produce. For example in volleyball three tasks were used: (a) the rebounding of a wall (wall drill), (b) having a ball thrown (actually lobbed) from one partner to another (fed drill), and (c) a 1 on 1 back and forth drill using set and bump. The data from Study 2 indicates the following.

1. The fed drill produced the lowest responding and the lowest error rates. It was a controlled environment for the student receiving the ball. Accuracy of response or “getting it right” was the goal. A mistake made in one response did not affect the next response significantly (other than the time taken to collect the ball by the throwing partner). Low skilled students made the biggest gains here, high skilled students immediately achieved high response rates for this task.

2. The wall drill produced very high rates of correct responding with proportional increases from baseline for incorrect responding (though the total rate of incorrect
responding was quite low). Responses here were part of a long behavior chain. Students who did not have accurate responses produced several errors and spent considerable time chasing the ball and having to restart. In this drill, making a mistake had a significantly impact on the total number of responses a student could make. If you were skilled the responses could be as high as 60 per minute. If a student took 10 seconds to retrieve a ball, then the total opportunities might be curtailed by as much as one-sixth (i.e., from 60 to 50 responses per minute. Were the student to make only 3 errors the total opportunities for responses might be reduced to below 30 per minute. The task focus was on fluency and endurance depending upon the time allowed for responding. Low skilled students did not do well in this drill, high skilled students did.

3. The 1 on 1 drill produced the most variability in opportunity to respond because each response was dependent upon the other players response which in turn was dependent upon the first student's previous response, and so on. The 1 on 1 drill was the most game-like. Less responses were produced here than in the wall drill. Error rates were higher than the fed drill but lower than the wall drill. The goal of the task was on developing fluency and endurance during application. Low skilled students did not do well in this drill while high skilled students did.

Given the rates of correct responding and incorrect responding produced by each task, a reasonable conclusion would
be to describe a relation between tasks and goals as a function of skill level. Such a relation for the volleyball task used in Study 2 is summarized in Table 21.

Table 21

The Relationship Between Tasks and Goals

<table>
<thead>
<tr>
<th>Goal</th>
<th>Fed Drill</th>
<th>Wall drill</th>
<th>1 V 1</th>
<th>other tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Endurance</td>
<td></td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
</tbody>
</table>

Table 21 is arranged so that the tasks can be cross-referenced against accuracy, fluency, endurance and application goals described in the precision teaching literature in Chapter 2. The conclusions of this present investigation suggest that the fed drill is a good drill to develop accuracy and error free learning, for low skilled students, but it is probably unnecessary for high skilled students to spend a lot of time performing this task. The wall drill provides a lot of responses for those students who can correctly perform the skill. Gains can be made quickly. This is probably not a drill for those who have not mastered the fed drill. The 1 on 1 drill places students
in a game-like environment, and allows for alternate responses between players. The 1 on 1 drill as an application task can also include fluency and endurance.

An analysis of tasks systems has not as yet analyzed the difficulty of the task, nor the appropriateness of the task relative to the skill of students. Categorizing tasks according to their goals as a function of the type of responding they produce, has significant implications for pedagogical-content knowledge including such issues as the validity of the task and the instructional goal of the task relative to game play.

Cooper, Heron and Heward (1987) distinguish between operants which are defined by conditions where the rate of response is under the control of the student (a free operant) such as dribbling, and conditions where the rate of response is "opportunity-bound" as in the case of a batting where a being thrown to you. A response (hitting the ball) can only be made when the pitcher provides and opportunity by throwing the ball. This difference in free and restricted operants influences the rate at which response can be made.

In this investigation dribbling tasks (in study 1), and the wall drill (study 2), produced extremely high rates of responding (at times exceeding 1 per second). Both of these tasks produced responses that were free operants. The other tasks in each study (passing in basketball, 1v1, and fed drills, in volleyball and badminton) produced considerable lower levels of responding, and were restricted operants.
One way of classifying tasks in future investigations might be in terms of their ability to produce restricted or free operants. If the goal of an instructional episode is to produce tasks that have a high rate of response, then the free-restricted operant distinction may provide a criteria for assisting in the selection of such tasks. The selection of a free or a restricted operant has implications in the selection of the unit of measurement. Cooper et al., (1987) suggest that rate of response is an appropriate unit of measurement for free operant responses, however they suggest that for restricted operants, percentage of total responses is the preferred unit of measurement. Cooper et al., (1987) argued that rate is sensitive to changes in free operant responding, and since restricted operants are opportunity-bound the more appropriate measure is one which measures the responses relative to the opportunities provided.

Recommendations

This was the first study that has experimentally analyzed responding in physical education with the purpose of increasing rates of response and measuring the effects on the qualitative dimensions of the response, as well as other variables including the nature of the task, and time use. Recommendations for future research are presented under the headings of methods, treatments, tasks, and games.
Methods of Analysis

The success of a response is important to obtaining a complete picture of the response. In this study the video camera had to be zoomed close enough to detect the topography of the response, and keep all students in view of the lens, but still capture the outcome of responses. This proved impossible using one camera and 4 target students. Rather than capturing an incomplete record of each, only measures on topography were used since these were for the most part, the focus of the teachers contingencies and instruction in the 3 studies. Future studies need to find ways to capture all the relevant dimensions of a response.

Procedural fidelity in this investigation was integral to stating conclusions about the effects of the dependent variable. Gresham, Gansle and Noell (1993) recently reported that in the Journal of Applied Behavior Analysis from 1980 to 1990 only 16% of the studies reviewed measured procedural fidelity. They commented “Failure to ensure that treatments are implemented as planned poses serious threats to the internal and external validity of experiments” (p. 257). In addition if our treatments are to be ecologically valid, researchers must measure the procedural fidelity of treatment implementation. Thus it is a recommendation of this study that measures of the procedural fidelity of the treatments, accompany measures of the dependent variables.
Treatments

Sprints and peers sprints certainly require replications across similar and different settings. The findings from this study require verification by other researchers, in particular questions about the effects of the treatments with tasks of varied difficulty and complexity require further analysis. For example it seems unlikely that a peer sprint would be useful in a game, or in modified games unless it was used in some form of group contingency.

The findings relative to hustles also require further examination. Particularly with respect to directing hustles and prompts to individuals rather than the group as whole. Much of what Tharp and Galimore cite in their paper as examples of John Wooden’s hustles are directed specifically at individuals, often with the effects generalizing to others, much like Kounin’s (1970) “ripple effect.” The choice of a fivefold increase from baseline was an arbitrary one. Perhaps effects that were equally as productive, but less demanding might have been achieved at a lower rate or by asking teachers to increase the rate of response to a more intense level without using a criterion. A second consideration in the hustles treatment is the issue of what occasioned the teacher’s behavior (increased responding and prompting). In the Tharp and Galimore study (1976) the goal of hustles was to increase levels of responding, and to correct errors. Player performance occasioned Wooden’s instructional behavior. In the case of this investigation,
teacher behavior was occasioned first by the researcher, teachers’ responded whether student behavior varied or didn’t vary since they were required to emit prompts and hustles at a particular rate for a period of time. Though there is little reason to suggest that the teachers did not respond to student errors and performance there is no evidence to support this contention.

Additionally, there were problems with the way the independent variable was operationalized. The unit of analysis was a single utterance. Typically teachers would repeat the prompt as in the example “go, go, go,” in this case the frequency of occurrence of the independent variable would be counted as 3 utterances. A more relevant analysis would be to treat the entire utterance as a phrase and count it as a frequency of one.

The influence of schedule effects also remains an important area to investigate. For example what would happen if the initial schedule of prompts were “thick” almost continuous reinforcement and then it “thinned” over time and was maintained by periodic “bursts”?

Among the more interesting outcomes of the use of the peer sprint condition was the nature of this treatment. Requiring students to discriminate and record the behavior of a partner appears at face value to be an important, though neglected line of inquiry. It remains unclear for this study what effect being able to discriminate another person’s behavior had
upon the student who did the discriminating, when that student’s turn to respond occurred.

Tasks

The tasks selected in this investigation with the exception of the badminton tasks were all commonly used by the teachers and were thus ecologically valid. Investigation into the effects of tasks commonly used by teachers, in producing success and their relevance to game play are important areas of future inquiry. It would appear logical that the content knowledge of the teachers (i.e., commonly used tasks) combined with treatments of the kind used in this investigation form a locus for research into pedagogical-content knowledge.

Games

Biscoff (1982) recommended using 3 male to 3 female ratios in physical education during game play to achieve equity of opportunity. The finding of this investigation would suggest that the size of the team and skill level are more important considerations in arranging instruction during games, rather than arranging teams by gender (as Biscoff suggested), when the goal is equity of opportunity. In this investigation no evidence of strategic play instruction occurred. In the games literature this is the most neglected area probably because it is the most neglected in games instruction.
APPENDIX A

COVER LETTER TO HIGH SCHOOL PRINCIPAL
The Principal
A Local High School
Columbus, Ohio 43207

Dear ,

I am beginning the dissertation phase of my Ph.D. program at The Ohio State University, School of Health, Physical Education, and Recreation. I would like to work with two teachers Mrs. East and Mrs. West from your school during my study. I have contacted both teachers and they have expressed a willingness to participate in this study. Therefore I am writing to you to obtain the necessary permission to conduct the study. If this is forthcoming I will then contact the district.

I have enclosed a brief description of the study, as well as an outline of the methodology, and the proposed time table.

Data collection is scheduled begin at the beginning of Winter quarter if approval is obtained from both yourself and the district. I would greatly appreciate your support in this endeavor.

Should you have any further questions, I would be delighted to answer them.

Sincerely,

Phillip Ward
APPENDIX B

PROSPECTUS
Title: An Experimental Analysis of Skill Responding in High School Physical Education

Investigators Name: Phillip Ward
Ph.D. Candidate

Address: School of Health, Physical Education, and Recreation
The Ohio State University
Room 315, Pomerene Hall
1760 Neil Ave
Columbus, OH 43210-1221.

Phone: 292-1292 (Office)
791-1119 (Home)

Graduate Advisor: Dr. Daryl Siedentop

I. Purpose of Study

The foci of this study is a comparison between three commonly used instructional strategies in physical education classes. The purpose of the study is to determine if these instructional strategies vary in their instructional effectiveness. Instructional effectiveness is defined in this study as student success (e.g., the number of correct/incorrect serves in a volleyball class). Along with student success rates, how students spend their time during the different instructional strategies will be observed (e.g., whether students are on task, off task, waiting, or assisting). This study will extend a twelve year research program at The Ohio State University which has investigated tasks and accountability in physical education. The research program is derived from and is similar to the research conducted in classrooms by Dr. Walter Doyle.
II. Methodology

A. Subjects for this study will be one secondary school physical education teacher and two of her grade seven classes. The classes will be observed for ten weeks while they engaged in their planned physical education activities.

B. Sessions will be audio and videotaped to help ensure the accuracy of data collection.

C. Data will be collected using systematic observation instruments designed to provide information about the type of tasks provided by teachers and the resultant performances by students (Lund, 1990); and information about how students spend their time during each task (Siedentop, Tousignant, & Parker, 1982).

D. Identification of teacher, students and schools will not be revealed in any publications or documents.

E. The teacher is not being asked to change her curriculum or her scheduled program of class activity. The study should in no way interrupt ongoing school activities.

F. No researcher will be required to have any contact with the students.

III. Time Schedule

A. The teacher serving as a subject has indicated a willingness to serve as a participant in this study.

B. It is hoped that when approval is obtained data collection may begin first day of class January 1993 and continue for the specified ten weeks.
Procedural Fidelity Checklist

Observer: _______________ Teacher: _______________

Date/Day: _______ Session #: ________ Activity: ________

Sprint

_____ Teacher called this condition a sprint

_____ Students paired

_____ One student performing other student is free

_____ Teacher asks students in class how many correct responses they did

_____ Teachers attention primarily focused upon the performing students

_____ 1 minute (+/- 5 seconds) timing

Peer Sprint

_____ Teacher called this condition a peer sprint

_____ Students paired

_____ Students received cards and pens

_____ One student performing other student observing/recording

_____ Teacher asked students who observed how many correct responses their partners did

_____ Teachers attention primarily focused upon the recording/observing students

_____ 1 minute (+/- 5 seconds) timing
REFERENCES


