This dissertation has been microfilmed exactly as received 69-22,116

DeBACY, Diane Lee, 1938-
The Effect of Viewing Videotapes Of a Selected Sport Skill Performed By Self and Others on Self-Assessment.

The Ohio State University, Ph.D., 1969
Education, physical

University Microfilms, Inc., Ann Arbor, Michigan
THE EFFECT OF VIEWING VIDEOTAPES OF A SELECTED
SPORT SKILL PERFORMED BY SELF AND
OTHERS ON SELF-ASSESSMENT

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

By
Diane Lee DeBacy, B.Ed., M.Ed.

The Ohio State University
1969

Approved by
Lewis A. Wes
Adviser
Department of Physical
Education
ACKNOWLEDGMENTS

I am indebted to those family and friends who have so willingly assisted in all aspects of this study. Without their patience and understanding this project would not have been completed.

Gratitude is expressed to my committee, Dr. Lewis A. Hess, Dr. Margaret A. Mordy and Dr. Chalmer G. Hixson for their helpful counsel and genuine interest in this dissertation.

I am also grateful to the Women's Physical Education Department at The Ohio State University for the use of their facilities and equipment. A special thanks is due to the women of the Physical Education staff who acted as judges and to those students who served as subjects in this study.
VITA

September 16, 1938 • Born - Lancaster, New York

1960 . . . . . . . . . . B.Ed., University of Buffalo
Buffalo, New York


1961-1966 . . . . . Instructor, Women's Physical Education Department, State University of New York at Buffalo, Buffalo, New York

1964 . . . . . . . . . . M.Ed., State University of New York at Buffalo, Buffalo, New York

1966-1967 . . . . . Teaching Assistant, Department of Physical Education for Women, The Ohio State University, Columbus, Ohio

1967-1968 . . . . . Assistant Professor, Department of Physical Education, State University of New York at Binghamton, Binghamton, New York

FIELDS OF STUDY

Major Field: Physical Education. Professors Margaret A. Mordy and Lewis A. Hess

Second Field: Teacher Preparation. Professor L. O. Andrews
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENT</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>Chapter</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Feedback and Self-Assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hypotheses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic Assumptions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limitations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Definitions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Significance of the Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Organization of Remaining Chapters</td>
<td></td>
</tr>
<tr>
<td>II.</td>
<td>BACKGROUND</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Perception and Judgment</td>
<td></td>
</tr>
<tr>
<td>III.</td>
<td>PROCEDURES</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Subjects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equipment and Facilities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Self-Assessment Scale</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment of Subjects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Judges' Assessments</td>
<td></td>
</tr>
<tr>
<td>IV.</td>
<td>RESULTS</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>General Considerations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test of Major Hypotheses: Accuracy of Self-Assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test of Minor Hypotheses: Change in Self-Assessment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test of Minor Hypotheses: Under- and Overassessment</td>
<td>iv</td>
</tr>
</tbody>
</table>
CONTENTS (contd.)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. DISCUSSION</td>
<td>53</td>
</tr>
<tr>
<td>General Considerations</td>
<td></td>
</tr>
<tr>
<td>Discussion of Major Hypotheses</td>
<td></td>
</tr>
<tr>
<td>Relevance of Data to Other Studies</td>
<td></td>
</tr>
<tr>
<td>Discussion of Minor Hypotheses</td>
<td></td>
</tr>
<tr>
<td>VI. SUMMARY AND CONCLUSIONS</td>
<td>76</td>
</tr>
<tr>
<td>Suggestions for Further Research</td>
<td></td>
</tr>
<tr>
<td>APPENDIX</td>
<td>81</td>
</tr>
<tr>
<td>A. Floor Plan.</td>
<td>81</td>
</tr>
<tr>
<td>B. Judges' Qualifications</td>
<td>82</td>
</tr>
<tr>
<td>C. Self-Assessment Rating Form and Instructions to Subjects</td>
<td>84</td>
</tr>
<tr>
<td>D. Score Values for Self-Assessment Scale.</td>
<td>88</td>
</tr>
<tr>
<td>E. Subjects' Orientation</td>
<td>89</td>
</tr>
<tr>
<td>F. Personal Data Card</td>
<td>91</td>
</tr>
<tr>
<td>G. Raw Data.</td>
<td>92</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>96</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Significance of the Difference Between Means of Pre Self-Assessment and Judges' Assessments</td>
<td>35</td>
</tr>
<tr>
<td>2. Judges' Mean Scores, Standard Deviations and Intercorrelations for all Subjects</td>
<td>36</td>
</tr>
<tr>
<td>3. Numbers Changing Toward and Away From the Mean Score of the Judges</td>
<td>38</td>
</tr>
<tr>
<td>4. A Summary of Change Toward Accuracy in Self-Assessment</td>
<td>39</td>
</tr>
<tr>
<td>5. Correlations Between Pre Self-Assessment, Post Self-Assessment and Judges' Assessment</td>
<td>40</td>
</tr>
<tr>
<td>6. Significance of the Difference Between Means of Post Self-Assessment and Judges' Assessment Scores</td>
<td>42</td>
</tr>
<tr>
<td>7. Means and Standard Deviations of Pre and Post Self-Assessment Scores for the Experimental and Control Groups</td>
<td>43</td>
</tr>
<tr>
<td>8. Analysis of Covariance for Post Self-Assessment</td>
<td>45</td>
</tr>
<tr>
<td>9. Adjusted Mean Post Self-Assessment Scores for the Experimental and Control Groups</td>
<td>45</td>
</tr>
<tr>
<td>10. Summary of Change in the Experimental and Control Groups</td>
<td>46</td>
</tr>
<tr>
<td>11. Number of Subjects Over- and Underassessing</td>
<td>48</td>
</tr>
<tr>
<td>12. Percentages, Proportions and t ratios for Subjects Overassessing Their Skill on Pre and/or Post Self-Assessment</td>
<td>49</td>
</tr>
<tr>
<td>13. Mean Pre and Post Overassessment Scores</td>
<td>51</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Vision is our most helpful sense; it is silent, takes up no room, and is capable of integrating complex information and providing aid with little interference. Electronic developments which enable one to view oneself performing and immediately correct actions seem to hold great promise for the future.¹

Videotape recorders, which provide optimal and immediate feedback of performance, are a relatively new phenomenon in physical education. The effective use of the videotape recorder constitutes a significant challenge for research because it is a dynamic medium being used without adequate research foundations.

There is a paucity of reported research related to the use of the videotape recorder in physical education. This, in light of 1) the growing popularity of video replay for instructional purposes; 2) the enthusiastic endorsement given it on the basis of little more than conjecture; 3) its obvious potential; and 4) the "insignificant difference trend" in the research findings, led the

writer to examine both the research and the instructional practices related to the use of the videotape recorder in physical education.

The investigator is aware of only five reported studies\(^2\) that involve videotape replay in physical education. All of these investigations were "relative effectiveness" studies in which the performance of students instructed with the aid of the videotape recorder was compared with the performance of others instructed without the aid of the videotape recorder. Of the five mentioned videotape studies, only Plese\(^3\) reported a significant difference in performance in favor of instruction utilizing


\(^3\)Plese, "Videotape Replay in Teaching Gymnastics."
videotape replay.

Deficiencies in experimental design were apparent in a number of these studies. The use of non-random groups, small samples and the confounding of variables were the most common faults. Speculative reasons for the failure of these studies to find significant differences focus on the discriminating ability of the measuring instruments used, the possibility of variable complexes negating each other and the concern for a single variable not significantly potent to produce significant differences.

It has become a rather prevalent practice in physical education to utilize the videotape recorder in conjunction with sport skill instruction. A basic pattern of instructional use, as evidenced through observation and reading by the writer, is typified as follows. Students receive instruction on the form to be used in executing a particular sport skill. This usually involves providing the students with a gross frame-work idea of the skill through demonstration or a movie. Students are expected to simulate the form presented, and, after some period of practice, they are taped while performing the skill. At some time later students view themselves performing via videotape replay. Viewing may be done individually or in groups, and critical comments recorded on the
tape or made personally by the instructor may or may not accompany the viewing.

The writer does not wish to imply that this is the only practice followed in physical education classes and acknowledges the fact that the procedure outlined represents a gross simplification of possible uses. However, it is not the writer’s purpose to present an extensive report on the use of the videotape recorder in physical education.

The use of the videotape recorder in this manner implies a very basic assumption. The assumption being made is that students viewing their performances via videotape replay perceive them accurately in terms of the skills involved. The question is how accurately does the student perceive his own performance, and what, in terms of perception change, is the nature of the effect of viewing one’s own performance on videotape? This study is directed toward these questions.

Chapter II presents the development of the rationale of the investigation. This was based upon an understanding of perception and judgment as it relates to viewing oneself and others performing a sport skill. While the perceptual process is the central focus of this study, a limited discussion of feedback and self-assessment as they contribute
to an understanding of this investigation follows.

**Feedback and Self-Assessment**

Feedback is information which an individual receives as a direct result of a response. The importance of feedback and/or knowledge of results has become a well known factor in learning. In order for the learner to judge his performance as he works for improvement, a reference or standard must be established by someone, e.g., the student or the teacher. The necessity of error information in guiding the acquisition of skill has been recognized by psychologists working in the human performance area.

The feedback of central importance to this study, that supplied by videotape replay, can be classified as augmented, terminal feedback—augmented because "an external source is providing feedback about the consequences of the movement" and terminal because it is given after a defined performance.

The existence or availability of feedback is no guarantee that an individual, even if he perceives it, does so with accuracy. Feedback consists merely of stimuli, and the feedback received through videotape replay

---


is subject to all the same factors that control and limit perception.

To the writer's knowledge there is no experimental data available related to videotape replay in the self-assessing or assessing of sport skills. In the area of self-assessment, the study most directly related to the present investigation was done by Brandt\(^6\) who used sixth and eleventh graders to study the accuracy of self-estimate. "The accuracy of self-estimate refers to the degree to which students correctly rate their own abilities and social reputations."\(^7\) Physical competency was one of two specific areas investigated and self-rating and performance data were obtained for broad jumping, strength of hand grip and baseball throwing. "The major finding of the research was that between-individual variation in accuracy of self-estimate was significantly greater than within-individual variation."\(^8\) Two additional findings pertinent to the proposed study were "performance ability was only slightly related to accuracy of self-estimate . . .


\(^7\)Ibid., p. 93.

\(^8\)Ibid., p. 93.
and self-rating accuracy may be developmental and tend to increase with age."\(^9\)

Though the present investigation was not concerned with self-perception, several conclusions drawn by Smith and Clifton based on their studies of self-perception as it relates to motor performance seem worthy of consideration. These conclusions are as follows.

1. Self-concept of motor performance will change when the perceptual field is expanded.

2. The highly skilled female performer viewing a motion picture of a 'model' performance and a motion picture of her own performance of the same movement patterns has no effect on her concept of her own performance.

3. College age females of low skill revise their concepts of their own performances in a negative direction after viewing motion pictures of a 'model' performance and of their own performances.\(^10\)

It should be noted that despite similarity of title, six major differences distinguish the present study from a particular investigation conducted by Smith

\(^9\)Ibid., p. 94.

\(^10\)Marguerite A. Clifton, "Mirrors of Movement," Purdue University, March, 1965. (Mimeographed)
1. The independent variable in the present study was viewing one's own performance as contrasted to the viewing of the model performance.

2. The present study was concerned with a self-assessment of skill based on observation as contrasted to an assessment based on responses made by the subject concerning her perceptions of her movements in performing motor skills.

3. Videotape replay was used to provide immediate feedback as contrasted to a four-week delay in viewing the loop films.

4. The present study was not only focused on change of assessment or perception, but also on the accuracy of assessment. This was not a consideration in the Smith and Clifton study; consequently, the reported directionalities of change could not be fully interpreted.

5. The present study provided four reference models for subjects to identify with as compared to one used in the former study.

On the basis of this study, Smith and Clifton con-

---

11Hope M. Smith and Marguerite A. Clifton, "Effect of Viewing Loop Films of One's Own Performance of Selected Movement Patterns and the Performance of a 'Model' on Self-assessment" (paper read at the American Association of Health, Physical Education and Recreation Conference, Dallas, Texas, March, 1964).
cluded that the self-assessments of the highly skilled were affected little or not at all and that lowly skilled subjects tended to re-assess their performances in a negative direction.12

STATEMENT OF THE PROBLEM

It was the purpose of this study to investigate the accuracy of self-assessment of a sport skill, the golf swing, by women students in beginning golf classes and to determine the effectiveness of videotape replay in reducing any differences between actual and self-assessed skill.

HYPOTHESES

The general hypothesis of this study is that viewing oneself performing a sport skill will reduce any difference between self-assessed skill and actual skill.

Although accuracy of self-assessment is the primary focus of this investigation, the writer felt that two areas of lesser significance were worthy of investigation. Minor hypotheses were tested in relation to the two secondary areas of study, change in self-assessment and the nature of the under- and overassessing tendency.

12Ibid.
In line with the general hypothesis, four major hypotheses were tested. These hypotheses, expressed in the null form, are:

I. That, in the experimental group, there is no difference between the number of self-assessments changed toward the mean score of the judges and the number of self-assessments changed away from the mean score of the judges.

It is predicted that a greater number of self-assessments will change toward the mean score of the judges than away from the mean.

II. That, in the control group, there is no difference between the number of self-assessments changed toward the mean score of the judges and the number of self-assessments changed away from the mean score of the judges.

The predicted direction of change for the control group is the same as that predicted for the experimental group.

III. That there is no difference between the mean post self-assessment and the mean score of the judges for the control group.

IV. That there is no difference between the mean post self-assessment and the mean score of the judges for the experimental group.
The null hypotheses tested regarding change in self-assessment were:

A. That there is no difference between the mean pre and post self-assessment scores for the control group.

B. That there is no difference between the mean pre and post self-assessment scores for the experimental group.

C. That there is no difference between the mean post self-assessment scores for the experimental and control groups.

D. That there is no difference between the mean change in self-assessment scores for the control and experimental groups.

The null hypotheses tested regarding under- and overassessment were:

E. That the pre self-assessments are evenly divided between the two conditions of under- and overassessment.

F. That the post self-assessments are evenly divided between the two conditions of under- and overassessment.

G. That there is no difference between the proportions of the experimental and control group showing a reduction in overassessment.
H. That there is no difference between the mean pre and post overassessment scores.

I. That there is no difference between the mean pre overassessment of the experimental and control groups.

J. That there is no difference between the mean post overassessment of the experimental and control groups.

BASIC ASSUMPTIONS UNDERLYING THIS STUDY

The following assumptions were inherent in the design of this study.

1. That individuals will honestly reveal self-assessments in a nonthreatening situation.

2. That the subjects made an honest effort in the performance of the physical task.

3. That the Physical Education Department's health records accurately reflected the visual status of the subjects.

4. That accuracy of self-assessment is an important variable in the learning of sport skills.

LIMITATIONS

This study does not attempt to investigate the factors which limit or influence perception, therefore, no inferences can be made regarding cause and effect of
these variables. Any inferences obtained from the results of this study are applicable only to the population sampled. The self-assessment scale used to secure all the data for this study was based on four model performances assumed to be representative of four skill levels approximating equal intervals on a skill continuum. It is recognized that there were limitations inherent in the construction of the scale. This study was further limited by the innate perceptual differences of the individual judges.

DEFINITIONS

Accuracy of self-assessment - refers to the degree to which students correctly assess their own sports skill levels

Assessment - "any procedure for making meaningful evaluations or differentiations among human beings with respect to any characteristic or attribute"\(^{13}\)

Form - the method of performance; "Performing within the confines of good mechanical principles . . . results in the most effective action, and it is this that we call

good form."\textsuperscript{14}

**Judge** - a person making an evaluation of any characteristic on the basis of any kind of behavior

**Model Performance** - a model representative of a predetermined level of skill; in this study, an exemplary performance of a golf swing

**Perception** - a complex process by which people select, organize and interpret sensory stimulations in association with previous experiences. We perceive a particular motor act such as a golf swing.

**Television Instant Replay** - the relatively immediate playback of a video recording

**SIGNIFICANCE OF THE STUDY**

The importance of feedback in performance and learning is universally agreed upon, but even when appropriate information exists, students obviously differ in their capabilities to perceive and utilize it. Though the videotape recorder has been used to provide students with feedback information, the investigator is not aware

of any attempt made to study the accuracy with which students perceive their own performances. The accuracy of students' perceptions pertaining to their own performances is basic to the learning process for these determine, to some extent, the error information derived from self-viewing and the goal-setting behavior of the individual.

The fact that the effect of viewing one's own performances via video replay has not been investigated on a more fundamental level with fewer operant variables than are found in a teaching-learning situation is indicative of what is unknown. This study, hopefully of heuristic value, was an attempt to arrive at a more fundamental level for investigating the effect of self-viewing as it possibly relates to both ongoing practices and past research.

ORGANIZATION OF THE REMAINING CHAPTERS

Chapter II presents some theoretical aspects of perception and judgment instrumental in the development of the rationale of the investigation. Chapter III gives an account of the procedures used in constructing the self-assessment scale and in collecting the data. The results of the investigation are reported in Chapter IV and discussed in Chapter V. Chapter V also presents a summary of the major findings and conclusions and contains suggestions for further research.
CHAPTER II

BACKGROUND

The obvious lack of allied research in the area of self-assessment as it relates to sport skills is indicative of the potential heuristic value of this study. The writer felt that any findings of this study would therefore be more meaningful and conducive to the generation of a questioning approach if they could be viewed from the broad perspective established in this chapter.

In this particular investigation, individuals were asked to make judgments of their own skill by means of a rather straightforward comparative process - one which required them to compare visually their skill to that of others by viewing video replays. However, when assessment is self-conducted, the validity or accuracy of an assessment may be limited by a multitude of variables. This particular discussion is limited to those aspects of perception and judgment instrumental in developing background for use in interpreting and questioning the results of this study.
PERCEPTION AND JUDGMENT

There would have been no need to conduct this investigation or like investigations if we were able to observe directly an individual's perceptual processes as we can the responses resulting from the processes. The observed responses are usually in the form of a judgment and are assumed to communicate the underlying perceptual processes.

Further complicating the study of perception is the fact that "... it is not easy to distinguish perceiving from its supervenient judgments."\(^{15}\) Perception and judgment can be differentiated theoretically, but in practical situations in which judgment supervenes on or expresses perceiving, it is difficult to separate the two processes.

Visual perception was of primary concern in the approach used to study self-assessment as affected by viewing videotape replays. Visual acuity was also a concern, but only to the extent that it did not negatively affect viewing the images on the monitors. To avoid confusion, it should be understood that visual perception

and visual acuity are two separate and distinct capacities. Visual acuity is the capacity to discriminate fine detail in the visual field while perception, as defined by Fellows, is "... the process by means of which an organism receives and analyses sensory information."\(^{16}\) By definition then, perception "... seems to divide into two operations: reception and analyses."\(^{17}\)

In visual perception, reception occurs when the retinas are exposed to the sensory input in the form of light energy. It is recognized that sensation, "... the immediate and direct apprehension of simple stimuli,"\(^{18}\) is primary to visual perception. However, it is important here only as it contributes to an understanding of the factors influencing perception. This relationship is summarized nicely by Berelson.

In short, sensory information does not correspond simply to the perception that it underlies. The fundamental reason for the difference between sensory data and perception is that sensory impulses do not act on an empty organism; they interact with certain predis-


\(^{17}\)Ibid.

positions and states already there, and immediate experience is the result of that interaction. The nature of the experience, then, depends on two interacting sets of contributions: those of the environment, in the form of physical stimulation, and those of the observer himself.\textsuperscript{19}

The definition of perception proposed by Fellows presupposes two operations: reception, a fairly well understood operation, and analysis, an operation far less understood. The analysis of sensory information seems to be further subdivided into two operations, most often labeled "attention" and "organization". Selection is the main characteristic of attention.\textsuperscript{20}

The myriad of sensory information that constantly bombards our sense receptors necessitates that we do not perceive indiscriminantly, but rather selectively. On a conscious level, an individual will look at certain aspects of stimulus material and ignore others. Beyond the direction chosen for one's eyes, the selection of stimuli that will become part of actual experience depends upon three major factors:

\begin{quote}
... the nature of the stimuli involved; previous experience or learning as it affects the observer's expectations (what he is prepared
\end{quote}

\textsuperscript{19}\textit{Ibid.}, pp. 99-100.
\textsuperscript{20}\textit{Fellows, Discrimination Process}, p. 5.
or set to see); and the motives in play at the time, . . . in short, what the observer wants or needs to see and not see.21

The selective process, according to Fellows, seems to have three effects. "That is, attending to a stimulus magnifies its influence. Second, attention also serves to reduce the influence of competing stimuli."22 "... attention also operates to isolate certain characteristics of the input."23 It should be noted that the selective process, as presented, does not involve any supplementation or transformation of the basic input. "... the information actually getting through is not changed by the process, but only reinforced or inhibited."24

Perception, as an area of study, is not free of controversies. This is reflected by the lack of agreement surrounding the organizational aspects of perception. It is obvious that the selection process, as interpreted, cannot account for all that is known about perception. In light of this it seems most appropriate to limit any explanation of the organization process to a general

21Berelson and Steiner, Human Behavior, p. 100.
22Fellows, Discrimination Process, p. 5.
23Ibid., p. 6.
24Ibid.
statement attributing any changes that occur in the information received as sensory input to the organizational process.

To avoid becoming entangled in a theoretical controversy, the present usage of judgment should be clarified. Consistent with the position taken in defining perception, judgment will be described as "... a conclusive or decisive process, not a productive one..."25 As stated earlier, perception may be pre-judicial and, if it is, the output generated by the perceptual process becomes the basis or material for judgment.

Obviously, judgments vary in difficulty. Johnson has classified judgments as being either complex or simple based upon the clarity of the stimulus aspect(s) to be judged.

By simple judgments we mean those in which the subject judges a simple, clearly specified aspect of the stimulus situation. . . . Complex judgments are those in which the object of judgment is complex and the stimulus or aspects to be judged are not distinctive.26


26 Ibid., p. 286.
Johnson, with support from Hollingworth\textsuperscript{27}, presents a rather good case for differentiating perception and judgment on a quantitative basis. He proposes that the distinguishing characteristic is the clarity of the perceived difference. Immediate perception is possible when a difference is clearly perceived, while judgment is exercised when differences are not clearly perceived. The difference between perception and judgment according to Johnson, then, is a quantitative one. "Presumably as confidence approaches a maximum and the time of response approaches a reaction time, judgment is minimized, and the act approaches direct perception."\textsuperscript{28}

If Johnson's position of differentiation is accepted as tenable, then a response to a task involving comparison and discrimination could be dominated by either perception or judgment. It would further follow that the "ease of judgment" experienced in carrying out such a task is probably dependent upon this quantitative factor. Consequently, tasks that do not involve using identical stimuli are subject to the influence of this variable, the quantitative factor.


Judgment, just as perception, has definite antecedents and consequences. Ideally, the preparation for perceptual interpretation or judgment establishes a set. "... as the perceptual set is broadened and becomes more complex and richly patterned with experience, the individual becomes capable of extracting more information from the environment." When "the stimulus material is heterogeneous, with no one prominent aspect or dimension to which the judge can be easily prepared to respond," experience will play a large part in the organization of the stimulus material. Certain aspects of a situation or object may stand out as figure to the experienced observer while the inexperienced observer's pattern of organization will be much less refined. An awareness of what is likely to be present in the perceptual field will facilitate differential selectivity, a characteristic distinguishing the trained observer from the untrained.


30 Johnson, Psychology of Thought, p. 293.

31 Figure-ground perception is the simplest pattern of stimuli organization. The figure is the focus of attention and all else in the visual field is relegated to ground or background status. A certain degree of heterogeneity within the visual field is essential for the perception of figure.
If the stimulus material is complex, the individual asked to make a judgment may find it necessary to prepare a more specific set by deliberately setting up the criteria for judgment. He may also isolate attributes he considers important as a basis for his response or he may even choose to rely upon a global impression rather than upon criteria that might result from analysis. All this implies that, even if the criteria to be used in making a judgment are specified, there is still a question as to which criteria were actually used in making an assessment.

It is apparent that the resulting judgmental response is dependent upon both the preparation for judgment and the stimulus material itself. Johnson points out that even though the preparatory set and the stimulus material are independent variables, they are not independent of each other. The set determines what is attended to while variation in the stimulus material facilitates the adoption and maintenance of some sets rather than others.

Obviously, such factors as objectivity, the de-

---

gree of satisfaction or dissatisfaction associated with one's own personal appearance, and impressions, reportable or unreportable, that were created in the past can affect one's self-assessment. However, though the amount of influence exerted by these factors is a matter of conjecture, their inclusion for discussion is beyond the scope of this chapter.
CHAPTER III

PROCEDURES

The primary purpose of this investigation was to determine the effect of viewing the golf swing performed by self and others on self-assessment. A secondary purpose of the investigation was to ascertain how accurately beginning golf students actually assess their own skill. In order to achieve these purposes, the following procedures were adopted.

SUBJECTS

Sixty-three women students enrolled in beginning golf classes at The Ohio State University participated in this study. Any woman with an uncorrected vision problem, according to the health records of the Physical Education Department, was not included in the study. Also, none of the subjects had ever seen their golf swing.

Participation in this study was voluntary, and subjects were solicited from eleven sections of Golf I classes taught by six different instructors. Members of the experimental and control groups were randomly selected and exposed to one of the two treatments individually.
Due to the availability of the facilities, equipment, and subjects two-thirds of both the experimental and control groups were exposed to the treatment after the seventh week of instruction and one-third was exposed after the eighth week of instruction.

**EQUIPMENT AND FACILITIES**

The videotape recording equipment used for this study consisted of a Sony VCK-2000 camera equipped with a Sony TV Zoom Lens (f=20-80mm), two Sony CV-2000 tape decks and an 18-inch Sony monitor. The tape duplication necessary to produce the sequence of four model performances, ordered from lowest to highest skill, was done on two Sony CV-2200 Videocorder Duplicators. The names used to identify the model performances were dubbed in during the tape duplication.

Two floodlights were used to supplement normal illumination. A diagram of the floor plan and equipment lay-out can be found in Appendix A.

This study was conducted at the indoor golf range in the women's field house at The Ohio State University, a facility consisting of nine hitting stations enclosed by a nylon net. Each station was equipped with a 5' x 4' rubber mat with a grass inlay. Live balls and number five
irons were provided for use in this study.

**SELF-ASSESSMENT SCALE**

Five judges, all experienced teachers of golf, (see Appendix B) selected four model performances to be used as reference points on the self-assessment scale designed for use in this study. The four model performances of a golf swing were representative of four levels of skill on a skill continuum. Selection was made from a pool of twenty-six performances, all videotaped in the same format used in taping the subjects. The judges were instructed to select performances that would approximate, as closely as possible, equal intervals on a skill continuum. They were further instructed that the low and high skill models selected should be such that subjects assessing themselves could conceivably fall outside of the extremes, i.e., judge themselves to be poorer than the lowest skill level depicted or better than the highest skill level depicted.

The self-assessment scale allowed for twelve possible judgments, each assumed to approximate equal intervals for scoring purposes. Self-assessment scores were assigned values ranging from one to twelve; one represented the lowest possible self-assessment and twelve
represented the highest. Subjects indicated if they judged themselves to be poorer than, better than or equal to the model performance they felt most closely approximated their own performance.

A copy of the self-assessment scale and the directions given to the subjects can be found in Appendix C. Appendix D contains a diagram of the scale with the scoring values indicated.

TREATMENT OF SUBJECTS

Each subject for this study was required to report for a single testing period at an assigned time and went through the entire procedure individually. Subjects first read a standardized statement (see Appendix E) which served three purposes. First, it briefly stated the purpose of the study. Secondly, it gave a limited account of the procedure to be followed. Thirdly, the statement explained the self-assessment scale used as a basis for all the scores. Subjects were asked if they had any questions about what they had read. When necessary the statement of explanation was supplemented by the investigator during the course of the taping and viewing. Next, a data card was completed by each subject (see Appendix F).
Subjects were allowed to warm-up with a five iron they had selected. The length of the warm-up period was determined by the individual and was conducted under the same lighting and physical conditions used for the taping.

Each subject hit four regular golf balls off the imitation grass inlay while being taped. An assistant placed each ball on the hitting surface. Subjects were instructed to step up to address the ball, hit it, step back and then repeat the procedure for the next ball.

Following the taping, each subject viewed the videotape of the four model performances and arrived at a self-assessment score by comparing her swing, as she mentally pictured it to be, to the models' golf swings. The self-assessment score was then recorded by the subject on the form provided (see Appendix C). All viewing was done on an 18-inch Sony Monitor from a movable chair. Subjects with corrective lenses were instructed to wear them while viewing the tapes at a comfortable distance from the monitor. All subjects, whether they fell into the experimental or control groups, followed the same procedure up to this point.

Subjects assigned to the control group then viewed the model performance tape a second time and were asked to record a second self-assessment on a duplicate
form. It was explained that the second viewing might or might not alter their original assessments.

Subjects assigned to the experimental group viewed their own performances immediately followed by a second viewing of the model performances. The experimental subjects were instructed to base their second and final self-assessment scores on a comparison between their swings and the model swings. The second self-assessment score was secured on a duplicate form.

JUDGES' ASSESSMENTS

The five judges that initially selected the four model performances were also used to rate each subject's golf swing. The initial meeting of the judges served as a training session since much discussion preceded the unanimous agreement on each model selected for use. In preparation for the rating of subjects, judges independently rated several performances not associated with the study and discussed any differences in ratings in an attempt to reach a satisfactory state of agreement on the criteria.

The original five judges viewed all the subjects' performances on videotape and assessed them on the same scale used by the subjects. Each judge independently recorded the assessment for each subject on a separate card.
The judges viewed the model performances before the rating session began and then again after rating every fifteen subjects. The mean score of the judges was accepted as the actual skill criterion score.

The procedure that was followed is outlined below.

**EXPERIMENTAL GROUP  N=31**

1. S warms up and hits four golf balls while being taped
2. S views the four model performances of the golf swing in ranked order (from low to high skill)
3. S records self-assessment score
4. S views her own performance and the model performances
5. S records a second self-assessment score
6. Judges rate each S's performance on the same scale used by the Ss

**CONTROL GROUP  N=32**

1. Same
2. Same
3. Same
4. S views only model performances
5. Same
6. Same
CHAPTER IV

RESULTS

General Considerations

This study employed a pretest-posttest control group design. The experimental and control groups were exposed to one of two treatments individually, with exposures of both types temporally and spatially intermixed.

The data in this study\textsuperscript{34} were analyzed in three classifications: an experimental group which consisted of thirty-one subjects; a control group which consisted of thirty-two subjects; and a combined group consisting of both the experimental and control groups.

Individuals were randomly assigned to the experimental and control groups so it was assumed that the groups were equal in terms of the independent variables, pre self-assessment and skill as reflected by the judges' assessments. Since the validity of this assumption is basic to the interpretation of the results of this study, an analysis of the pre self-assessments and judges' as-

\textsuperscript{34}The raw data upon which this investigation was based are reported in their entirety in Appendix G.
sessments was made to insure that the two groups were not significantly different in respect to these variables. The null hypotheses tested were:

1. That there is no difference between the mean scores of the judges for the control and the experimental group.

2. That there is no difference between the mean pre self-assessments for the control and the experimental group.

The statistic used to test the hypotheses was the t-test for uncorrelated data.35 Throughout the data analysis, whenever appropriate, an F-test36 was made to test the null hypothesis of no difference between the variances before an estimate of the standard error of the mean difference was computed. Table 1 presents the results of the analyses used to test the significance of the mean differences.

To test the hypotheses, a two-sided test and a significance level of .05 were chosen. On the basis of the obtained values of t, .434 and 1.31, neither hypothesis

could be rejected. Therefore, it can be assumed with some certainty that the two randomly selected groups were not, in fact, significantly different in skill or in their pre self-assessments. Even though neither of the null hypotheses was rejected, it is interesting to note that the difference between the mean pre score and the mean judges' score was significant at the .05 level in both the experimental and control groups.

**TABLE 1**

**SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS OF PRE SELF-ASSESSMENT AND JUDGES' ASSESSMENT SCORES**

<table>
<thead>
<tr>
<th></th>
<th>Experimental Group (N=31)</th>
<th>Control Group (N=32)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean pre self-assessment score</td>
<td>4.806</td>
<td>5.469</td>
<td>1.310</td>
</tr>
<tr>
<td>Variance</td>
<td>3.960</td>
<td>3.806</td>
<td></td>
</tr>
<tr>
<td>Mean judges' score</td>
<td>2.955</td>
<td>2.759</td>
<td>-0.434</td>
</tr>
<tr>
<td>Variance</td>
<td>3.201</td>
<td>2.983</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>2.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.92&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>significant at the .05 level

The judges' assessment scores were accepted as the skill criterion measure. A summary of the individual
judge's mean scores and standard deviations as well as the intercorrelations of the five judges is presented in Table 2.

### TABLE 2

**JUDGES' MEAN SCORES, STANDARD DEVIATIONS AND INTERCORRELATIONS FOR ALL SUBJECTS**

<table>
<thead>
<tr>
<th>Judges</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean score</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.714</td>
<td>1.641</td>
</tr>
<tr>
<td>2</td>
<td>0.639</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.667</td>
<td>1.492</td>
</tr>
<tr>
<td>3</td>
<td>0.778</td>
<td>0.645</td>
<td></td>
<td></td>
<td></td>
<td>2.841</td>
<td>2.391</td>
</tr>
<tr>
<td>4</td>
<td>0.796</td>
<td>0.744</td>
<td>0.800</td>
<td></td>
<td></td>
<td>3.111</td>
<td>2.025</td>
</tr>
<tr>
<td>5</td>
<td>0.761</td>
<td>0.738</td>
<td>0.795</td>
<td>0.750</td>
<td></td>
<td>2.921</td>
<td>2.217</td>
</tr>
</tbody>
</table>

The mutual relationship between judges are represented by the intercorrelations. It can be observed that only two of the ten reported correlations were less than .7. The correlation of Judge 2 with Judge 1 and with Judge 3 was .6 < .7.

**Test of Major Hypotheses: Accuracy of Self-Assessment**

In order to investigate the accuracy of self-assessment, it was necessary to examine the pre and post
scores of a subject relative to a skill criterion measure. The mean score of the judges was accepted as being representative of an individual’s skill.

An examination of the direction of change in self-assessment relative to the mean score of the judges indicates that a majority of the subjects in the experimental and control groups combined changed toward more accurate self-assessments. The two hypotheses that were tested relative to the direction of change were:

**Hypothesis I** - That in the experimental group there is no difference between the number of self-assessments changed toward the mean score of the judges and the number of self-assessments changed away from the mean score of the judges.

**Hypothesis II** - That in the control group there is no difference between the number of self-assessments changed toward the mean score of the judges and the number of self-assessments changed away from the mean score of the judges.

The sign test\(^7\) for two related samples provided the test for Hypotheses I and II, which state that the median of the difference is zero. The cases used in the computation were limited to those showing a change—twenty-three in the experimental group and twelve in the

control group. Both hypotheses were tested using the .05 level of significance and an upper-tailed region for rejection since a change toward the mean score of the judges was predicted.

Table 3 summarizes the direction of change data and the results of the sign tests for both groups.

**TABLE 3**

<table>
<thead>
<tr>
<th>NUMBERS CHANGING TOWARD AND AWAY FROM THE MEAN SCORE OF THE JUDGES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Experimental (N=31)</td>
</tr>
<tr>
<td>Control (N=32)</td>
</tr>
</tbody>
</table>

For \( P \geq .5 \), the proportion under the hypothesis tested, a \( k \geq 16 \), the number of cases changing toward the mean score of the judges, was required to reject the null hypothesis pertaining to the experimental group. Table 3 indicates \( k = 16 \), so Hypothesis I was rejected.

A \( k \geq 10 \) was needed to reject Hypothesis II.

Table 3 indicates only four members of the control group changed toward the mean score of the judges—a number insufficient to cause rejection of the hypothesis.
Clearly, the direction of change in self-assessment for the experimental group was in the predicted direction, thus being indicative of a "gain in accuracy". To calculate this gain for each subject that changed toward the mean score of the judges, the following formula was used:

\[
\text{gain in accuracy} = \frac{\text{pre score} - \text{judges' mean}}{\text{post score} - \text{judges' mean}}
\]

A summary of the "gain in accuracy" data located in Table 4 will be discussed in Chapter V.

**TABLE 4**

**A SUMMARY OF CHANGE TOWARD ACCURACY IN SELF-ASSESSMENT**

<table>
<thead>
<tr>
<th>Group</th>
<th>Number Changing Toward Accuracy</th>
<th>Total Change</th>
<th>Mean Change^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>16</td>
<td>30</td>
<td>.97</td>
</tr>
<tr>
<td>(N=31)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>4</td>
<td>5</td>
<td>.16</td>
</tr>
<tr>
<td>(N=32)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^aMean changes were based on N's = 31 and 32 cases.

Some indication of accuracy of self-assessment may be gained by testing the significance of the correlations between 1) the judges' score, accepted as being repre-
sentative of actual skill, and pre self-assessment; and
2) the judges' score and the post self-assessment for
both groups. A further analysis of between-group dif­
ference was made by using Fisher's Z-transformation to
test the significance of the difference between the two
sample correlation coefficients. The correlations are
presented and compared in Table 5.

| TABLE 5 |
|------------------|------------------|
| CORRELATIONS BETWEEN PRE SELF-ASSESSMENT, POST |
| SELF-ASSESSMENT AND JUDGES' ASSESSMENT |
| Assessment | Variables | Experimental | Control | |
| | | Group (N=31) | Group (N=32) | |
| | | r | p | r | p | Z |
| Pre vs post | 0.656 | .01 | 0.900 | .001 |
| Pre vs judges' | 0.361 | .05 | 0.244 | NS | .49 |
| Post vs judges' | 0.427 | .05 | 0.251 | NS | .74 |

An inspection of Table 5 shows that all three of
the correlations reported for the experimental group were
significant beyond the .05 level. The only significant
correlation for the control group was that between the

38 Downie and Heath, Basic Statistical, p. 144.
pre and post scores.

Fisher's Z-transformation was used to test the significance of the difference between the correlations of the experimental and control groups for post score with the judges' score. The obtained Z value, considerably less than 1.96, indicated that the difference was not a significant one.

Finally, an analysis of the difference between the mean post self-assessment and the mean score of the judges was made. The significance of the mean difference was tested by means of the t-test. The level of acceptance was pre-set at .05.

The two null hypotheses tested were:

Hypothesis III - That there is no difference between the mean post self-assessment and the mean score of the judges for the control group.

Hypothesis IV - That there is no difference between the mean post self-assessment and the mean score of the judges for the experimental group.

The data used to test Hypotheses III and IV and the results of the t-tests are located in Table 6.

On the basis of the available information, Hypothesis III was rejected at the .05 level. However, the mean difference in the experimental group was not sufficient to cause rejection of Hypothesis IV.
**TABLE 6**

SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS OF POST SELF-ASSESSMENT AND JUDGES' ASSESSMENT SCORES

<table>
<thead>
<tr>
<th></th>
<th>Mean Post Self-Assessment</th>
<th>S.D.</th>
<th>Mean Judges' Assessment</th>
<th>S.D.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>group N=31</td>
<td>4.13</td>
<td>2.93</td>
<td>2.95</td>
<td>1.79</td>
<td>1.90</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>group N=32</td>
<td>5.69</td>
<td>2.07</td>
<td>2.76</td>
<td>1.73</td>
<td>6.23a</td>
</tr>
</tbody>
</table>

*a* significant at the .05 level

**Test of Minor Hypotheses: Change in Self-Assessment**

In this section, the change under consideration is limited to the difference between the pre and post self-assessments. The changes in self-assessment will be examined in terms of within-group differences and between-group differences. Accuracy, or change in relation to the judges' assessment scores, was considered in the preceding section.

In order to determine whether any significant within-group changes in self-assessment occurred, the significance of the differences between means was tested by the t-test. The .05 level of confidence was selected as
the criterion for rejection of the null hypotheses.

The null hypotheses tested were:

Hypothesis A - That there is no difference between the mean pre and post self-assessment scores for the control group.

Hypothesis B - That there is no difference between the mean pre and post self-assessment scores for the experimental group.

The results of the analyses are presented in Table 7.

| TABLE 7 |
|---|---|---|---|
| MEANS AND STANDARD DEVIATIONS OF THE PRE AND POST SELF-ASSESSMENT SCORES FOR THE EXPERIMENTAL AND CONTROL GROUPS |

<table>
<thead>
<tr>
<th>Groups</th>
<th>Pre Self-Assessment</th>
<th>Post Self-Assessment</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Experimental</td>
<td>4.806</td>
<td>1.990</td>
<td>4.129</td>
</tr>
<tr>
<td>(N=31)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>5.469</td>
<td>1.951</td>
<td>5.688</td>
</tr>
<tr>
<td>(N=32)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No significant within-group difference was found for either the control or experimental group. Therefore, Hypotheses A and B cannot be rejected, and it must be concluded that the within-group change, the difference be-
tween pre and post self-assessment, is not significant at the .05 level for either group.

A further analysis was made of the difference between the experimental and control treatment by means of analysis of covariance. Analysis of covariance was used to make allowances for possible skill differences between the two groups. The judges' scores, accepted as being representative of actual skill, were designated as the covariate (i.e., the source of variance to be eliminated through the statistical procedure). The analysis of covariance table (Table 8) presents information regarding the degree of freedom, sum of squares and mean squares derived from the analysis. The adjusted post score means and standard errors are presented in Table 9.

The null hypothesis tested was:

Hypothesis C - That there is no difference between the mean post self-assessment scores for the experimental and control groups.

The obtained F ratio, \( F = 4.543, \text{ df } = 1,60 \), was significant at the .05 level and led to the rejection of Hypothesis C. It appears then that after adjusting with a covariate, there was a significant difference between the two treatments.

In order to provide a more complete analysis of change, the significance of the difference between the
### TABLE 8

ANALYSIS OF COVARIANCE FOR POST SELF-ASSESSMENT

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum-Squares (due)</th>
<th>Sum-Squares (about)</th>
<th>df</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (between)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (within)</td>
<td>61</td>
<td>218.4848</td>
<td>171.8741</td>
<td>60</td>
<td>2.8646</td>
</tr>
<tr>
<td>Treatment and error (total)</td>
<td>62</td>
<td>243.7143</td>
<td>184.8889</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Difference for testing adjusted means</td>
<td></td>
<td></td>
<td>13.0148</td>
<td>1</td>
<td>13.0148</td>
</tr>
</tbody>
</table>

### TABLE 9

ADJUSTED MEAN POST SELF-ASSESSMENT SCORES FOR THE EXPERIMENTAL AND CONTROL GROUPS

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Post-Score Mean</th>
<th>Adjusted Mean</th>
<th>SE Adjusted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>4.1290</td>
<td>4.4522</td>
<td>0.3062</td>
</tr>
<tr>
<td>Control</td>
<td>5.6875</td>
<td>5.3745</td>
<td>0.3013</td>
</tr>
</tbody>
</table>
mean changes in the two groups was tested by computing a z-score.\textsuperscript{39}

The null hypothesis tested was:

**Hypothesis D** — That there is no difference between the mean change in self-assessment scores for the control and experimental groups.

Table 10 presents 1) a summary of the data related to change (change being defined as the difference between the pre and post scores irrespective of the direction); and 2) the z-score computed to test the null hypothesis.

**TABLE 10**

**SUMMARY OF CHANGE IN THE EXPERIMENTAL AND THE CONTROL GROUPS**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Total Change</th>
<th>Mean Change</th>
<th>Variance'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>55</td>
<td>1.774</td>
<td>2.046</td>
</tr>
<tr>
<td>(N=31)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>17</td>
<td>0.531</td>
<td>0.561</td>
</tr>
<tr>
<td>(N=32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z value</td>
<td></td>
<td>4.23\textsuperscript{a}</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Significant beyond the .01 level

\textsuperscript{39}Ibid., p. 136.
Hypothesis D was rejected since the obtained z value of 4.23 is significant beyond the .01 level.

Test of Minor Hypotheses: Under- and Overassessing Tendency

The data were examined for evidence of a trend by subjects to overassess their skill. A subject was considered to have overassessed her skill when her self-assessment score was higher than the mean score assigned her by the judges. In order to determine if such a trend did exist in either or both groups, data concerning the frequency of overassessment by subjects were first subjected to the sign test.40

The null hypotheses tested for both the experimental and control groups were:

Hypothesis E - That the pre self-assessments are evenly divided between the two conditions of under- and overassessment.

Hypothesis F - That the post self-assessments are evenly divided between the two conditions of under- and overassessment.

Both hypotheses were tested for each group using a .05 upper-tailed region of rejection since overassessment was predicted. Table 11 presents a summary of the numbers of subjects under- and overassessing on the pre

---

and post scores and the results of the four sign tests.

### TABLE 11

**NUMBER OF SUBJECTS OVER- AND UNDERASSESSING**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Number Overassessing</th>
<th>Number Underassessing</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experimental</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre score</td>
<td>23</td>
<td>8</td>
<td>.05</td>
</tr>
<tr>
<td>post score</td>
<td>15</td>
<td>14</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre score</td>
<td>28</td>
<td>4</td>
<td>.05</td>
</tr>
<tr>
<td>post score</td>
<td>28</td>
<td>3</td>
<td>.05</td>
</tr>
</tbody>
</table>

To examine the overassessment tendency more thoroughly, the data were divided into five categories for comparison. A summary of the proportions and the per cent equivalents for each category are presented in Table 12. The t ratios for testing the significance of the difference between two proportions are also reported in Table 12.

An inspection of Table 12 indicates that significant differences at the .01 level were found between the proportions in categories 2, 3 and 4. However, no sig-
### Table 12

**PERCENTAGES, PROPORTIONS AND t RATIOS FOR SUBJECTS OVERASSESSING THEIR SKILL ON PRE AND/OR POST SELF-ASSESSMENTS**

<table>
<thead>
<tr>
<th>Category</th>
<th>Control</th>
<th>Experimental</th>
<th>Difference $^b$</th>
<th>t value $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>N$^a$</td>
<td>Percent</td>
<td>N$^a$</td>
</tr>
<tr>
<td>1. Pre self-assessment</td>
<td>87.5</td>
<td>28/32$^d$</td>
<td>74.2</td>
<td>23/31$^d$</td>
</tr>
<tr>
<td>2. Post self-assessment</td>
<td>87.5</td>
<td>28/32$^d$</td>
<td>48.4</td>
<td>15/31</td>
</tr>
<tr>
<td>3. Pre and post self-assessment</td>
<td>84.4</td>
<td>27/32$^d$</td>
<td>45.2</td>
<td>14/31</td>
</tr>
<tr>
<td>5. Post but not pre self-assessment</td>
<td>3.6</td>
<td>1/28</td>
<td>6.7</td>
<td>1/15</td>
</tr>
</tbody>
</table>

$^a$Proportion of subjects in each cell

$^b$Difference = Per cent of control group minus per cent of experimental group

$^c$t value for testing the null hypothesis of no difference between two proportions

$^d$Ho: $P \leq .50$ rejected for reported proportions

$^e$Significant at the .01 level with df = $(N_1-1) + (N_2-1)$
significant difference was found for categories 1 and 5.

A further analysis of the data was made concerning those subjects who overassessed their skill in both the pre and post assessments. In the experimental group, seven out of fourteen, or 50 per cent of the subjects who overassessed on both the pre and post assessments did so to a lesser degree on the post score. The comparable statistics for the control group reveal that three out of twenty-seven, or 11.1 per cent of the subjects reduced their overassessments on the post scores.

A t-test was employed to test the significance of the difference between these two measures. The null hypothesis tested was:

Hypothesis G - That there is no difference between the proportions of the experimental and control groups showing a reduction in overassessment.

The obtained value of $t$, ($t = 2.79$), was significant at the .01 level and led to the rejection of Hypothesis G.

The final analysis in this section concerns the actual magnitude of overassessment. The mean pre and post overassessment scores were compared to determine the significance of between- and within-group differences.

The null hypothesis used to test the within-group differences in both the experimental and control groups
was:

Hypothesis H - That there is no difference between the mean pre and post overassessment scores.

The null hypotheses applied to test the between-group differences were:

Hypothesis I - That there is no difference between the mean pre overassessment of the experimental and control groups.

Hypothesis J - That there is no difference between the mean post overassessment of the experimental and control groups.

The mean overassessment data and results of the t-tests used to test the above null hypotheses can be found in Table 13.

**TABLE 13**

**MEAN PRE AND POST OVERASSESSMENT SCORES**

<table>
<thead>
<tr>
<th></th>
<th>Experimental Group (N=31)</th>
<th>Control Group (N=32)</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Over-assessment</td>
<td>SD</td>
<td>Mean Over-assessment</td>
</tr>
<tr>
<td>Pre self-assessment</td>
<td>2.08</td>
<td>1.80</td>
<td>2.94</td>
</tr>
<tr>
<td>Post self-assessment</td>
<td>1.62</td>
<td>2.33</td>
<td>3.13</td>
</tr>
<tr>
<td>t value</td>
<td>0.92</td>
<td></td>
<td>0.44</td>
</tr>
</tbody>
</table>

a=significant at the .05 level
Neither of the within-group differences tested under Hypothesis H was significant at the .05 level. The results of testing Hypothesis I revealed that the difference between the pre overassessment means was not significant. The only significant difference found was that between the mean post overassessments. Thus, Hypothesis J was rejected at the .05 level.
CHAPTER V

DISCUSSION

Four major hypotheses and ten minor hypotheses were tested. The major hypotheses focused on the accuracy of self-assessment while the minor hypotheses dealt with change in self-assessment and the under- and over-assessment tendency in self-assessment. The organization of the present chapter will parallel that of the preceding chapter in regard to the order of areas discussed.

General Considerations

All subjects used in this study 1) were women students enrolled in beginning golf classes at The Ohio State University; 2) were, according to the Physical Education Department's health records, free of visual defects that would interfere with viewing a TV monitor; and 3) had never seen their golf swings. These criteria were used to control the variables of similarity of golf experience, sex and vision. Subjects were randomly assigned to either the experimental or control treatment, but, to insure that systematic bias had not been introduced into the groups by the assignment procedure, a t-test was per-
formed on the difference between the mean pre-scores of the two groups and on the difference between the mean score of the judges for each group. The hypotheses which predicted that there would be no differences between the mean pre-score of the two groups or between the mean judges' scores for the groups were supported. Thus, it can be concluded that the assignment procedure did not lead to systematic biasing of the groups.

An inspection of the mean scores (Table 1, p. 35) indicates that, even though the control group was slightly less skilled than the experimental group according to the judges, their pre-score mean was slightly higher than that of the experimental group.

In light of the procedures followed by the judges, the investigator did not feel it was imperative to determine statistically the intra-judge reliability. The selection of the model performances for the rating scale entailed the viewing, independent rating, discussion and re-viewing of approximately thirty performances. Viewing, rating and discussion were continued until there was unanimous agreement on the models to be used.

Following selection of the model performances, the judges assessed performances on the basis of the rating
scale under the same viewing conditions which were to be used in the actual study. This procedure, accompanied by discussion, was continued until the judges displayed consistent agreement.

The agreement of the judges under the actual experimental conditions was determined by the correlation technique. Although lower than anticipated, the intra-judge correlations reported in Table 2 on page 36 were considered to be satisfactory. One factor which may have contributed to the lower-than-anticipated correlations is the possible existence of true disagreement on the part of the judges regarding the golf swing. Although the preparation for rating the subjects' swings involved much discussion, it was impossible to ascertain the relative importance the judges placed on the particular components of the swing or swing errors. However, the existence of such differences in opinion are not uncommon, even among experts.

At an early stage in the planning of the study, it was anticipated that the sample drawn for the investigation would be representative of a range of skills comparable to the skill range illustrated by the model performances. This would have allowed for testing between-skill levels to investigate the relationship between
performance level and accuracy of self-assessment.

However, it should be noted that the mean score of the judges for the total sample studied (N = 63) was only 2.86. This indicated that the sample drawn was a predominantly low skilled group. Although the average level of skill displayed by the subjects was higher than that of the poorest model, but lower than that of the second poorest model, their skill was actually closer to that of the poorest model.

There may be several reasons for the unexpected high frequency of low skill subjects that volunteered for this study. First, it is logical to find the distribution in beginning golf classes more concentrated in the low-skill category. Therefore, even a random sample of beginning golf classes would likely contain an exceptionally large percentage of students who would fall below the mid-point of the scale.

Secondly, it is quite possible that the less skilled students were greatly motivated to participate in this study. The motivation may have stemmed from the fact that this study provided these students with an opportunity that they judged to be potentially helpful in improving their skill—namely viewing themselves on
videotape.\textsuperscript{41}

In short, the number of subjects falling into the upper half of the skill range was negligible, thus making it impossible to test any hypotheses dealing with skill level differences.

**Discussion of Major Hypotheses: Accuracy of Self-Assessment**

The results of testing Hypothesis I, the null hypothesis of no difference between the number of self-assessments changed toward and away from the judges' mean score, indicated that the experimental treatment resulted in a change toward more accurate self-assessment in a proportion of the subjects, \( P \leq .5 \), significant at the .05 level.

The number of subjects in the control group that changed toward a more accurate assessment was equivalent to only one-third of the total number showing a change. Obviously, the control treatment did not improve the accuracy in a significant proportion of the control subjects.

\textsuperscript{41}Potential subjects were told that participation in this study would afford them the opportunity to view themselves on videotape. Control subjects saw a replay of their swings at the completion of their participation in the study.
To extend the analysis of accuracy one step further, each group's mean "gain in accuracy" (Table 4, p.39) was compared to its mean post overassessment (Table 13, p. 51). The mean post overassessment for the experimental group (N=31) was 1.62 as compared to a mean "gain in accuracy" of .97. The "gain" in the experimental group then can be interpreted as approximately a 37 per cent improvement in accuracy or, in other words, a 37 per cent reduction in overassessment.

In the control group the mean post overassessment was 3.13 while the mean "gain in accuracy" was only .16. In this case, the "gain" represents an improvement in accuracy of only approximately 5 per cent.

The fact that some members of the control group did show a change when given the opportunity to re-assess their skill might seem unusual since viewing the model performances a second time was all that intervened between assessments. The actual magnitude of the change was not unreasonably large—eight of the twelve subjects showing a change, changed only one scale unit each, while the other four subjects changed two units each.

One possible explanation for change in the control group might be that it was a manifestation of the Hawthorne Effect. Another tenable explanation for the observed change
centers about the possible apprehensions and expectancies or lack of expectancies associated with doing something for the first time. Details might have been attended to with finer discrimination the second time the model performances were viewed, thus, producing a change in the original assessment.

The results of testing the final major hypotheses, III and IV, revealed that the difference between the post self-assessment mean and the mean of the judges' was significant only in the case of the control group.

It will be recalled that the difference between the pre-assessment mean and the mean score of the judges proved to be significant for both the experimental and control groups (Table 1, p. 35). Consequently, the fact that the mean difference between the post self-assessments and the judges' score for the experimental group was not significant leads to the tentative conclusion that the experimental treatment was effective in producing a statistically significant improvement in accuracy.

Although not directly related to any of the major hypotheses, a test of the difference between the correlation coefficients of the judges' scores with the post self-assessments for the two groups was made. The difference was not significant, but an inspection of Table 5
on page 40 shows that the correlation coefficient associated with the experimental group was significant. However, this significance is indicative of a very weak relationship. The correlation of the control group (r = .251) suggests that skill and the final self-assessments were not related for this group.

From the analyses of the data presented in conjunction with the testing and discussion of the four major hypotheses, the following conclusions were drawn:

1. The experimental treatment produced a gain in accuracy in a proportion (P ≤ .5) of the experimental subjects significant at the .05 level.

2. The control treatment did not improve the accuracy of assessment in a proportion (P ≤ .5) of the subjects significant at the .05 level.

3. The experimental treatment was effective in improving the accuracy of self-assessment.

Thus, it was concluded that the data lent support to the general hypothesis that viewing oneself performing a sport skill will reduce any difference between self-assessed skill and actual skill.

Thus far, the discussion has focused on the effects of the experimental and control treatments on accuracy of self-assessment. Results supported the premise that self-
viewing would improve accuracy, but a considerable dis-
parity is still apparent between actual and self-assessed
skill, even after viewing. Unfortunately, it is impos-
sible to explain the persisting inaccuracy on the basis
of the accumulated data or published research, but possi-
ble contributory factors can be discussed.

One complex of factors which conceivably could
have contributed to the failure of the experimental treat-
ment to produce a greater gain in accuracy of assessment
concerns the stimulus material. The judgmental responses
were based on viewing video replays of the golf swing, a
rather complex movement sequence. The comparative process
required for arriving at accurate assessments may have
exceeded the subjects' capacity for perceiving, thus,
preventing them from observing and remembering the impor-
tant aspects for judgment. This possibility was considered
in developing the experimental design of this study. How-
ever, a review of the research most closely paralleling
the present study revealed that Clifton and Smith con-
ducted a pilot study\textsuperscript{42} to determine the mean number of

\textsuperscript{42}M. A. Clifton and H. M. Smith, "Pilot Studies
in Perception of Oneself in Movement." Unpublished Stu-
dies, University of California, Los Angeles, 1958-1959.
viewings requested when students were encouraged to view moving pictures of themselves for as long as they wished. They found that four viewings was the mean number requested. This fact, coupled with the investigator's interest in approximating a typical class exposure, led to the adoption of four repetitions of the golf swing for viewing purposes. Possibly, unlimited viewing would have improved the accuracy of assessment.

Also related to a consideration of the stimulus material is the possible influence exerted on judgment by variables in the stimulus material. This possibility is further complicated by the fact that self-viewing introduced a variation in the stimulus material. As stated previously, (Chapter 2, p. 24), it is apparent that judgmental response is dependent upon both the stimulus material and the preparation for judgment. It is the preparation for judgment on which the discussion will now focus for its possible contributions to the inaccuracy displayed by the students.

To insure a high degree of accuracy in self-assessment, the preparatory set should involve much practice in identifying the important aspects of the object of judgment. In this particular investigation no attempt was made to instill a common set in the subjects to serve
as a basis or frame for judgment. This was avoided for two reasons. First, it was felt that this would contribute to the confounding of variables since a learning situation would have been introduced. Secondly, each subject brought to the experimental situation a preparatory set which evolved as a result of the golf instruction and experience each had prior to taking part in the study.

At best, the preparation for judging a motor skill could be inadequate for making accurate comparative judgments such as those called for in this study. The complexity of the judgment is often compounded when subjects fail to adhere to criteria, even when they are specified. Obviously, only research designed to compare response to stimulus can identify just which aspects of the stimulus material a subject takes into account and which he ignores. Johnson has extracted several general principles of judgment from research of this kind relevant to this discussion.

1. The judgment may be influenced by stimulus aspects or variables to which attention is not directed by the explicit instructions or by the logical implications of the stimulus material.

2. The judgment may be influenced by stimulus variables which the judge can not or does not report.
3. When the judgment called for is avoided, because of its difficulty or for any other reason, judgment of some other stimulus aspect will be made.43

All of this implies that the actual criteria subjects used for judgment probably varied considerably from making separate judgments on parts of the whole golf swing—probably an action or actions emphasized by the instructor—to relying upon a general impression of the swings viewed. This general impression may have been due to some very striking aspect of an action which caused other aspects of the action to be ignored or dismissed as being of minor importance. This practice is likely to result in inaccuracy for often times the individual fails to realize that he is operating with this type of set.

In response to the question: "Can you mentally picture a good golf swing?", (data card, Appendix F) thirty-seven, or approximately 80 per cent, of the experimental subjects replied "yes". Even so, there still exists the possibility that viewing the model performances may have modified their pre-determined mental set, hence, the criteria used for judgment.

From the foregoing discussion, it appears quite

43Donald Johnson, Thought and Judgment, pp. 294-296.
tenable that the observed disparity between actual and self-assessed skill may, in fact, be attributed to the nature of the stimulus material and the preparation for judgment or the lack of it.

Relevance of Data to Other Studies

A number of "relative effectiveness" studies were referred to in the introduction of this paper, the results of which were, with one exception, non-significant (Caine, 1966; Hirsch, 1968; Lebrecque, 1968; Penman et al., 1968; and Plese, 1967).

As was pointed out in the first section of the paper, a main criticism of these studies was the confounding of variables. The considerable number of operant variables detracted from the meaningfulness of the conclusions. It will be recalled that this was one reason the writer proposed to investigate the effect of viewing oneself on a more fundamental level than that used in the above mentioned studies.

The generalizations derived from this study only have some general validity to the degree that the experimental situations are typical of other situations to which one would like to generalize. Since the present study focused on one element both common and basic to all of the "relative effectiveness" studies mentioned—viewing
oneself performing for the purpose of evaluating a performance or some aspect of it—it seems less imprudent to hypothesize about the "no significant difference" findings associated with the majority of these studies.

The importance of feedback in learning has been well established, but little or no research has been done regarding the conditions which affect the interpretation of feedback provided by video replays of sport skills. How should viewing be conducted to be most effective?

In the "relative effectiveness" studies researched, subjects were asked to interpret the feedback provided by replays under a variety of conditions. It could well be that the "no significant difference" results were, in part, a function of the specific conditions accompanying viewing. Naturally, if subjects were unable to interpret the feedback accurately, the mere availability of error information would not result in any performance improvement.

The findings of this investigation strongly suggest that beginning students in a sport activity do not have

44 Margaret Robb, "Feedback".
the ability to accurately interpret the feedback provided by video replay when viewed under the conditions typified in this study. This implies that the potential of video replay as an adjunct to learning may not have been fully realized in the experimental situations.

Another study which has relevance for this dissertation was conducted by Smith and Clifton. The purpose of their study was "to determine if viewing motion pictures of one's own performance of selected movement patterns along with films of a 'model' performance of the same patterns causes a change in one's self-assessment of his own performance." Subjects for the study were thirty-six physical education majors and thirty-three non-majors.

Smith and Clifton concluded, as a result of their investigation, that "persons of low skill level, after viewing films of their own performance, tend to re-assess their own performance in a negative direction." The present study provides evidence for the support of this conclusion and further suggests that the re-assessments might have represented more accurate appraisals.

---

45Smith and Clifton, "Effect of Viewing Loop Films."
46Ibid.
Discussion of Minor Hypotheses: Change in Self-Assessment

To answer the question, "Will there be any significant change in self-assessment due to self-viewing?", four minor hypotheses were tested. The null hypotheses tested dealt with three aspects of change: 1) the change represented by the difference between the pre-score mean and the post-score mean; 2) the end result of the change, as represented by the post-scores; and 3) the mean change, the difference between the pre- and post-scores irrespective of sign.

Rejection of null hypotheses A and B, that there are no within-group differences between the mean pre and post self-assessment scores, is clearly indicated in the case of both the experimental and the control groups. It is concluded, therefore, that neither treatment produced a significant amount of change. However, it is notable that the changes showed by the two groups were opposite in direction. The experimental treatment resulted in a mean decrease while the control treatment produced a mean increase in change. The direction of change will be more appropriately discussed in the next section which deals with under- and overassessment.
Hypothesis C dealt with the difference between the mean post self-assessments of the two groups. The analysis of covariance data signified that there was a significant difference between the means adjusted for skill. Even though the two groups appeared to be initially equivalent in skill, the existence of a slight mean difference in skill as represented by the mean score of the judges made it desirable to designate the judges' scores as the covariate and exclude the possible effects of skill difference in post self-assessment.

An examination of the adjusted post score means (Table 9, p. 45) shows that the effect of the analysis was to reduce the mean difference by approximately 59 per cent.

In light of the fact that there was no difference between the two groups in the pre self-assessments, the significance of the mean difference leads to the tentative conclusion that the experimental treatment produced significantly more change in assessment than did the control treatment.

The final analysis evaluated the net change produced by the two treatments. Hypothesis D was designed to test the significance of the difference between the changes made by the two groups, i.e., the mean change in
the experimental group minus the mean change in the control group. In terms of net change, the experimental treatment produced a mean change of 1.774 compared to a change of .531 for the control group. When this difference was tested, the analysis revealed that this was a significant difference at the .01 level.

In general, the data seem to indicate that self-viewing, as it was conducted in the experimental treatment, really did not cause a significant change in assessment. However, in comparing the change that did occur in the two groups, the experimental treatment appears to have resulted in significantly more change.

Discussion of Minor Hypotheses: Under- and Overassessing Tendency

The null hypothesis (E) of no difference between the two conditions of under- and overassessment in pre score was rejected for both the control and experimental groups. Proportions of both groups, significant at the .05 level, initially overassessed their skill.

Although both groups were similar in this respect, the results of testing Hypothesis F indicated that only members of the control group persisted in overassessing on the post self-assessment in numbers sufficient to cause rejection of the null hypothesis. As a consequence of the
experimental treatment, the number of subjects under-assessing on the post-score, fourteen closely approximates a balance with the fifteen subjects who overassessed.

The conclusions which may be drawn from testing Hypotheses E and F are:

1. In both groups a proportion of the subjects \((P \leq .5)\), significant at the .05 level, overassessed their skill on the pre self-assessment.

2. The control treatment did not significantly reduce the number of individuals who overassessed on the post-score from the number who initially overassessed their skill.

3. The experimental treatment significantly reduced the number of individuals who overassessed on the post-score from the number who initially overassessed their skill.

A more detailed examination of under- and over-assessing was undertaken by testing the difference between the proportions of corresponding categories for the experimental and control groups. On the pre self-assessment, (Table 12, Category 1, p. 49), there was no significant difference between the proportion of subjects overassessing in each of the two groups. This was evidence that the two groups were comparable in terms of numbers displaying
overassessment before each group was exposed to one of the two treatments.

A test of the proportions under categories 2, 3, and 4 resulted in additional support for the conclusion that self-viewing was effective in significantly reducing the number of subjects who overassessed on the pre-score.

The significant difference associated with category 2 confirmed and further substantiated the finding that the two groups did differ significantly in the number overassessing on the post self-assessment. This is due to the fact that the proportions tested under category 2 were based on the total number in each group while this was not necessarily the case in the proportions tested under Hypothesis F.

A comparison of the proportions in category 4 lends even more support to this conclusion since overassessment on the pre-score was a prerequisite for inclusion in this category. As might have been suspected, the difference between the proportions of those overassessing on both the pre and post assessments was significant.

The proportions reported in category 5 are simply a reflection of the discrimination shown in category 3. In connection with this category, it is notable that only one subject in the experimental group that did not
overassess on the pre-score did so on the post-score.

In short, the results of testing the difference between the proportions were consistent with and supportive of the findings which have evolved from testing Hypotheses E and F.

Hypothesis G focuses more closely on the assessments of those subjects who overassessed on both the pre and post-scores (category 3, Table 12, p. 49). Even though both appraisals by subjects represented an overassessment, the data analysis revealed that the re-assessments of seven out of fourteen, or 50 per cent of the experimental subjects in this category showed a reduction in the amount by which they overassessed. In comparison, only three out of twenty-seven members of the control group, or approximately 11.1 per cent, showed a reduction in overassessment when they re-assessed their skill.

The analyses of the under- and overassessing tendency up to this point have been concerned with the predominance of the overassessment tendency and the effect self-viewing has had on the numbers exhibiting this tendency. Obviously, self-viewing did result in a significant reduction of the number of subjects who overassessed. However, the last three hypotheses were designed to determine if the actual reductions produced were significant.
A cursory examination of the mean pre and post overassessments for each group (Table 13, p. 51) shows that both of the means associated with the control group exceeded those of the experimental group. Within-groups, the experimental treatment caused a reduction in overassessment, but not sufficiently large to be statistically significant. Likewise, the difference for the control group was insignificant. The absence of within-group differences is consistent with the failure of either treatment to produce significant within-group change (Table 7, p. 43).

A between-group comparison of the mean pre overassessment scores showed that the control mean was not significantly different than the experimental mean. The only significant difference found within or between the groups in the analyses of mean overassessment was that found between the mean post overassessment scores for the two groups. Thus, it appears that while neither of the treatments caused a significant within-group reduction in overassessment, the experimental treatment resulted in a reduction sufficient to cause a significant difference between the magnitudes of overassessment in the control and experimental groups.

Because of the relatedness of accuracy and over-
assessment, many of the findings related to under- and overassessment were predictable. However, this may not have been the case if the inaccuracies of the large majority of the subjects had not been in the direction of overassessment.

Due to the sample studied and the area investigated, only a general comment can be made comparing the overassessing tendency exhibited by subjects in this study to findings of other research. It appears that this tendency parallels that found in "level of aspiration" studies, i.e., poor performers tended to overrate themselves.
CHAPTER VI

SUMMARY AND CONCLUSIONS

This study was designed to obtain and analyze self-assessment data of women students enrolled in beginning golf and to explore several questions regarding the nature of self-assessment. It was focused particularly on the accuracy of self-assessment which refers to the degree of agreement between self-assessed skill and actual skill. The judges' ratings were accepted as being representative of actual skill. The nature of the changes in self-assessment and of the under- and over-assessing tendency were of secondary concern in this study.

Two self-assessments were secured from all subjects. The subjects viewed the four model performances which served as the basis for the rating scale developed for use in this study. The first, or pre self-assessment, was arrived at by having subjects compare their swings, as they pictured them to be, to those of the models. Control subjects repeated this procedure in arriving at their second or post self-assessments. Experimental subjects arrived at their post-assessments after viewing a replay of their swings followed by viewing the model performances. Judges later viewed a video replay of each subject's
swing and assessed it on the same scale used by the subjects.

Four major hypotheses were tested in connection with the general hypothesis that viewing oneself performing a sport skill will reduce any difference between self-assessed skill and actual skill. Minor hypotheses tested dealt with change in self-assessment and the nature of under- and overassessment.

The major finding of the research was that self-viewing did improve the accuracy of self-assessment.

Additional findings were:

1. Women students in beginning golf tended to hold inaccurate concepts of their actual skill as represented by their golf swings.
2. The self-assessments of women students in beginning golf tended to be inaccurate in the direction of overassessment.
3. The viewing of model performances alone did not improve the accuracy of self-assessment.
4. The viewing of model performances and one’s own performance produced a reduction in overassessment.

Suggestions for Further Research

Relatively little systematic research has been con-
ducted regarding the use of videotape replay in physical education. Evidence from previous studies as well as from the present one indicate the need for a sophisticated approach to research in this area.

Future research designs might focus on a considerable number of variables in different combinations for the purpose of studying interactions among variables. This would allow for the investigation of the effects various conditions have on interpretation and use of feedback provided by video replay.

Another area needing the attention of further research is the stimulus material itself. How should it be structured and organized to result in optimal gains in performance?

An extension of the present study might involve an in-depth study on the accuracy of self-assessment compared to the accuracy of assessing others. It is further suggested that research be done to determine the relationship between skill and accuracy of self-assessment. The present study failed to secure a sample with adequate representation of all skill levels, thus, it was impossible to investigate the highly skilled.

This study has numerous implications for the teaching-learning process. One suggestion for further
study involves the effects of various teaching methods on the accuracy of self-assessment. If viewing video replays is to be a prevalent practice in physical education, it behooves us to investigate the ways in which it can be best used. Should critical comment proceed, follow or accompany self-viewing to aid students in interpreting the feedback they receive from video replays? Does viewing model performances in conjunction with self-viewing confuse or enlighten students? These fundamental questions need to be researched.

Another area worthy of consideration concerns assessment by physical education majors. It would seem important to investigate the accuracy with which majors assess themselves and others. Knowledge of this kind would have implications for professional preparation.

Although under- and overassessment was not an area of major concern in this investigation, it is a topic worthy of major research. An exploration of the nature of under- and overassessment would contribute to the understanding of learning.

A very fertile area for further research focuses on the effect of self-viewing on the individual. What is the relationship between self-concept and accuracy of self-assessment? The self-concept has become a major area of
concern in education, and it would appear that it is of paramount importance in physical education.
APPENDIX A

FLOOR PLAN

Golf Mat

Target Area Was Located On Far Side Of Mat

10'

9'

55°

Camera
APPENDIX B

JUDGES' QUALIFICATIONS

Judge #1 taught physical education for eight years and taught golf on the college level for seven of the eight years. This person taught golf to beginning, intermediate and advanced students of golf. Judge #1 played golf on a recreational, club and tournament basis.

Judge #2 taught physical education for eight years, but taught golf for approximately twelve years. This person taught golf in public school, college, adult education and at a golf club. She taught golf to beginning, intermediate and advanced students of golf. Judge #2 played golf on a recreational, club and tournament basis.

Judge #3 taught physical education for thirty-two years and taught golf in a university school and college for twenty-eight of the thirty-two years. This person taught golf to beginning, intermediate and advanced students of golf. Judge #3 played golf on a recreational and club basis.
Judge #4 taught physical education for seventeen years and taught golf on the college level for nine of the seventeen years. This person taught golf to beginning, intermediate and advanced students of golf. Judge #4 played golf on a recreational, club and tournament basis.

Judge #5 taught physical education for thirteen years and taught golf on the college level and in adult education for eight of the thirteen years. This person taught golf to beginning and intermediate students of golf. Judge #5 played golf on a recreational basis.
APPENDIX C

SELF-ASSESSMENT RATING FORM
INSTRUCTIONS TO ALL SUBJECTS
INSTRUCTIONS TO ALL CONTROL SUBJECTS
INSTRUCTIONS TO ALL EXPERIMENTAL SUBJECTS
SELF-ASSESSMENT RATING FORM

KAY    BEV    DOT    SUE
INSTRUCTIONS GIVEN TO ALL SUBJECTS PRIOR TO VIEWING MODEL PERFORMANCES

"You will first view the four model performances of the golf swing. Each model you see will be identified by name and is represented on this self-assessment scale." (The subjects are given a self-assessment form.)

"The models will be seen in an order of increasing skill, with Kay, the poorest of the four models, being seen first. At the completion of the tape, you will be asked to draw a line through the scale at the point which best approximates your self-assessment of your golf swing. This assessment will be based upon your swing, as you picture it, compared to that of the models just seen. Remember, you may decide that your swing is poorer than, equal to or better than any one of the four models. However, if you feel your swing is located somewhere between two models, you must choose the model which more closely represents your skill. Do you have any questions?"

INSTRUCTIONS GIVEN TO CONTROL SUBJECTS PRIOR TO VIEWING MODEL PERFORMANCES THE SECOND TIME

"Now, I would like you to view the same model per-
formances a second time, and again indicate how you assess or rate your swing in comparison to the models. After viewing the models a second time, you may either repeat your original assessment if you feel it was accurate, or you may want to change it. Record your second self-assessment on this card." (First card is collected from subjects as second card is distributed.)

"Do you have any questions?"

INSTRUCTIONS GIVEN TO EXPERIMENTAL SUBJECTS PRIOR TO VIEWING THEMSELVES AND THE MODEL PERFORMANCES A SECOND TIME

"Now you will see a replay of your golf swing immediately followed by another replay of the model performances. Your first assessment was based upon a comparison between the way you imagined your golf swing to look and the way the model performances looked to you. On this card, (First card is collected from subjects as second card is distributed.), I would like you to rate or assess your swing as you actually see it in comparison to that of the models. Do you have any questions?"
APPENDIX D

SCORE VALUES FOR SELF-ASSESSMENT SCALE

<table>
<thead>
<tr>
<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>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAY</td>
<td>BEV</td>
<td>DOT</td>
<td>SUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

— © — i — © — 1 — e —
APPENDIX E

SUBJECTS' ORIENTATION

The purpose of the study you have volunteered to participate in is to learn more about self-assessment as it relates to a motor skill. Specifically, this study will try to ascertain how accurately you can actually assess your own skill in golf.

PROCEDURE - After warming up you will be videotaped while hitting four golf balls. You will then view four women's golf swings which are representative of four different skill levels, ranging from high to low. These four model performances will be used as a basis for your self-assessments.

SELF-ASSESSMENT SCALE EXPLANATION AND DIRECTIONS - Each of the four performances you will see are identified both on the scale pictured below and on the videotapes by name.
You will see the performances in an order of increasing skill going from the poorest performance by KAY located on the left end of the scale through to the best performance located on the right end of the scale.

After watching each performer take four swings you will draw a line through some point on the scale to indicate how you assess your skill in comparison to that of the model performances viewed. The divisions between model performances on the scale allow you to indicate which of two models more closely approximates your skill level. For example, the sample score marked on the scale indicates the person felt she was better than BEV, poorer than DOT, but that her own skill was really closer to that of BEV's. If you feel that one of the model performances itself best represents your skill level draw a line through the circle directly above the name.
Circle a response for each of the following

1. Year in school  Frosh  Soph  Jr  Sr
2. Did you play golf before taking golf at O.S.U.?  yes  no
3. Is this your first quarter of golf at O.S.U.?  yes  no
4. Have you ever seen your golf swing?  yes  no
5. Can you mentally picture a good golf swing?  yes  no
6. Which one of the following sources has contributed most to the picture you have of your golf swing?
   a. verbal comments  b. imitations by  c. information by the instructor  d. feelings or sensations from your own body
### APPENDIX G

#### TABLE 1

**RAW DATA: PRE AND POST SELF-ASSESSMENT SCORES AND INDIVIDUAL JUDGES' SCORES FOR THE EXPERIMENTAL GROUP**

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Pre Score</th>
<th>Post Score</th>
<th>Judges' Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
TABLE 1 (Contd)

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Pre Score</th>
<th>Post Score</th>
<th>Judges' Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>26</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>28</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>29</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>31</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Subjects</td>
<td>Pre Score</td>
<td>Post Score</td>
<td>Judges' Scores</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Subjects</td>
<td>Pre Score</td>
<td>Post Score</td>
<td>Judges' Scores</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>----------------</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>23</td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>27</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>28</td>
<td>7</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>31</td>
<td>6</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>
SELECTED BIBLIOGRAPHY

BOOKS


PERIODICALS


PROCEEDINGS, REPORTS AND YEARBOOKS


UNPUBLISHED STUDIES


