Increasing the accuracy of self-report data through the use of reliability enhancement and self-monitoring training procedures

Scott Brown Hamilton

The University of Montana

Follow this and additional works at: https://scholarworks.umt.edu/etd

Let us know how access to this document benefits you.

Recommended Citation


https://scholarworks.umt.edu/etd/5916

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
INCREASING THE ACCURACY OF SELF-REPORT DATA THROUGH THE USE OF RELIABILITY ENHANCEMENT AND SELF-MONITORING TRAINING PROCEDURES

By

Scott B. Hamilton
B.S., Colorado State University, 1973

Presented in partial fulfillment of the requirements for the degree of

Master of Arts
UNIVERSITY OF MONTANA
1976

Approved by:

[Signatures]

Chairman, Board of Examiners

Dean, Graduate School

Date

Nov. 3, 1976
The purpose of the present investigation was to: (a) assess the separate and combined effects of Self-monitoring Training (SMT) and Reliability Enhancement Package (REP) procedures on the self-monitoring accuracy of speech anxious undergraduates, (b) compare SMT to a procedure designed to control for exposure to irrelevant training stimuli (i.e., Training Control - TC) and REP to a manipulation composed of accuracy instructions and a simple "record" cue (i.e., Accuracy Instructions - AI), and (c) evaluate the effectiveness of experimental procedures when the demand for behavior change was either "low" (unmanipulated) or "high" (manipulated).

Twenty-eight speech anxious undergraduates, who had indicated an interest in participating in a speech anxiety treatment program, served as subjects. Seven subjects were included in each factorial combination (i.e., SMT-REP, SMT-AI, TC-REP, TC-AI) and all 28 participated in the three self-monitoring speech sessions (i.e., SM I, II, and III). Participants were initially informed that treatment would consist of repeated exposure (i.e., flooding) to the actual feared situation (i.e., public speaking) and that they would be responsible for evaluating their progress by self-monitoring objective signs of anxiety (i.e., frequency of speech disfluencies) as well as subjective perceptions of fear during each speech performance. Subjects self-monitored objective and subjective behaviors during the three separate speech sessions. The three speeches given during each session were unobtrusively recorded, thus allowing for a comparison to be made between the number of disfluencies self-monitored and the number actually emitted. Experimental procedures were administered between SM I and SM II and demand for behavior change was manipulated prior to the final speech session.

Results indicated that: (a) SMT was more effective than TC in increasing the accuracy of self-monitored data, (b) REP as a sole accuracy-enhancing manipulation was more effective than AI only when the target behavior was easily discriminable, and (c) a combination of SMT and REP procedures proved most effective in improving the accuracy of self-monitored data and in maintaining these gains over time.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
</tbody>
</table>

**Chapter**

<table>
<thead>
<tr>
<th>I</th>
<th>INTRODUCTION</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>External Observation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Self-monitoring</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>The Present Investigation</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>METHOD</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subjects</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Experimental Design</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Therapists</td>
<td>34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXPERIMENTAL PHASES</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-monitoring I (Session 1)</td>
<td>35</td>
</tr>
<tr>
<td>Instruction and Training (Session 2)</td>
<td>42</td>
</tr>
<tr>
<td>Reliability enhancement package (REP)</td>
<td>42</td>
</tr>
<tr>
<td>Accuracy instructions (AI)</td>
<td>42</td>
</tr>
<tr>
<td>Self-monitoring training (SMT)</td>
<td>42</td>
</tr>
<tr>
<td>Training control (TC)</td>
<td>42</td>
</tr>
<tr>
<td>Self-monitoring II and Demand Manipulation (Session 3)</td>
<td>49</td>
</tr>
<tr>
<td>Self-monitoring III (Session 4)</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TAPE RATER AND DEPENDENT MEASURES</th>
<th>50</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>RESULTS</th>
<th>54</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject Loss</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Inter-rater Reliability</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Change in Number of Self-monitoring Errors</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>&quot;Ah&quot; disfluencies</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Repetition disfluencies</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Combined disfluencies</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Change in Number of Disfluencies Emitted</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>&quot;Ah&quot; disfluencies</td>
<td>61</td>
</tr>
</tbody>
</table>
Repetition disfluencies
Combined disfluencies
Percentage of "Accurately" Monitored Speeches ........................................ 67
Correlations Between Self-recorders and Criterion Rater ............................. 71
Anxiety Reduction Measures ................................................................. 73
PRCS questionnaire
Mean Likert-scale ratings
Post-experimental Questionnaire ......................................................... 74

IV DISCUSSION ......................................................................................... 78

Self-monitoring Training ............................................................... 80
Reliability Enhancement Package ................................................... 81
Procedural Combinations ............................................................... 84
Therapeutic Effects ............................................................................ 87
Methodological Implications ............................................................ 90

V SUMMARY .......................................................................................... 97

REFERENCE NOTES ............................................................................. 100

REFERENCES ......................................................................................... 103

APPENDIX A: PERSONAL REPORT OF CONFIDENCE AS A SPEAKER (PRCS) ................................................................. 115
APPENDIX B: DATA SHEET I: OBJECTIVE ANXIETY ........................................ 118
APPENDIX C: DATA SHEET II: SUBJECTIVE ANXIETY .................................. 120
APPENDIX D: DATA SHEET SUMMARY .................................................. 122
APPENDIX E: SPEECH TOPICS AND SUBTOPICS ..................................... 124
APPENDIX F: SMT SPEECH TRANSCRIPTS ............................................... 128
APPENDIX G: SPEECH RATING FORM ....................................................... 134
APPENDIX H: PROGRAM EVALUATION QUESTIONNAIRE .......................... 136
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Experimental Design</td>
<td>33</td>
</tr>
<tr>
<td>2.</td>
<td>Summary of External-monitoring Training Procedure</td>
<td>47</td>
</tr>
<tr>
<td>3.</td>
<td>Change in Mean Number of Self-monitoring Errors From Pre-manipulation Period</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>(SM I)</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Change in Mean Number of Externally-monitored Disfluencies From Pre-</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>manipulation Period (SM I)</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Percentage of Observations (speeches) With Perfect or One Disfluency</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Deviation Between Self-recorders and Criterion Rater</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Mean Correlations Between the Number of Self- and Externally-monitored</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Speech Disfluencies (&quot;Ah's&quot; and Repetitions)</td>
<td></td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Approximate number of self-monitoring errors predicted for the four experimental conditions across the three self-monitoring periods</td>
<td>31</td>
</tr>
<tr>
<td>2.</td>
<td>Change in mean number of self-monitoring errors (for &quot;Ah's&quot; and Repetitions combined) from pre-manipulation period (SM I)</td>
<td>62</td>
</tr>
<tr>
<td>3.</td>
<td>Change in mean number of externally-monitored disfluencies (&quot;Ah's&quot; and Repetitions combined) from pre-manipulation period (SM I)</td>
<td>68</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

External Observation

Perhaps the most significant contribution made by the field of behavior therapy during the past decade has been in the development and refinement of procedures aimed at evaluating the effects of treatment interventions in applied settings (Lipinski & Nelson, 1974a; Kazdin, 1975a). The surge of interest in the assessment area has not only resulted in the development of creative single-subject designs (see Leitenberg, 1973), factorially sound self-report inventories (e.g., Tasto, Hickson, & Rubin, 1971), refined physiological measures (e.g., Bancroft, 1971), and innovative data collection devices (e.g., Grimaldi & Lichtenstein, 1969), but has also stimulated research on the most basic of the applied assessment strategies, i.e., behavioral observation.

Naturalistic behavioral observation has been considered to be the least inferential approach to treatment evaluation and has been widely advocated as a primary assessment tool for use in diverse clinical settings (Bushell, Wrobel, & Michaelis, 1968; Goldfried & Kent, 1972; Werry & Quay, 1969). Direct behavioral observation not only promotes
objective appraisal of symptom severity and possible maintaining factors (i.e., antecedents and consequences) but also leads to an ongoing assessment of behavior change during and following treatment administration. Data obtained from systematic behavioral observations allows the therapist to gain continual feedback regarding the effectiveness of the particular treatment strategies employed and permits application of necessary corrections in order to obtain maximal therapeutic benefits (Franks & Wilson, 1975). Unless the data collected by the external observer(s) is reliable and consistent, however, therapists may unwittingly modify treatment procedures in a non-therapeutic direction or may falsely assume that their intervention has had the desired therapeutic impact. Indeed, the usefulness of behavioral observation as an assessment device largely depends upon the ability of the observer(s) to collect accurate and unbiased data (Taplin & Reid, 1973).

At first glance, the problem of observer reliability would appear relatively straightforward and easily managed. In order to obtain reliable behavioral observations, one must simply operationally define terms and train observers to accurately monitor and record the occurrence of target behaviors prior to initiating actual data collection. However, in view of recent experimental findings, naive assumptions regarding the simplicity of the observational process have been shattered and researchers have been
forced to deal with a number of important and highly complex issues.

The first of these problems involves the actual behavior change of subjects precipitated by the presence of an external observer. For example, Zegiob, Arnold, and Forehand (1975) recently demonstrated that mothers, after being informed that their behavior was being observed, played significantly more with their children, were more positive in their verbal interactions, and structured play activities to a greater degree than during an uninformed observation period. The reactive effects of observer presence have also been noted in the behavior of family members (Patterson & Harris, Note 1; Johnson & Lobitz, Note 2; White, Note 3) nursery school children (Arsenian, 1943), teachers and students (Mercatoris & Craighead, 1974), as well as museum visitors (Bechtel, 1967). Although recent reviews in this area (Johnson & Bolstad, 1973; Wiggins, 1973) have suggested that reactivity can be minimized through the incorporation of prolonged adaptation periods (i.e., to allow subjects to habituate to the presence of an external observer), unobtrusive observational procedures (cf., Webb, Campbell, Schwartz, & Sechrest, 1966) have become increasingly popular in controlling for observer presence (e.g., Bornstein, Hamilton, Miller, Quevillon, & Spitzform, Note 4; Nelson, Lipinski, & Black, 1975; Surratt, Ulrich, & Hawkins, 1969).
A second frequently encountered problem in external-observation involves the dual concepts of instrument decay (Campbell & Stanley, 1966) and observer drift (O'Leary & Kent, 1972). Although it has been repeatedly assumed that observers will maintain the high level of accuracy demonstrated in initial training sessions throughout the entire data collection period, recent investigations have failed to support this assumption. Reid (1970) found a median drop of 25 percentage points in observer accuracy between the final day of overt reliability assessment (i.e., observers were informed that their accuracy was being checked) and the first day of covert (i.e., uninformed) assessment. Similarly, Romanczyk, Kent, Diament, and O'Leary (1973) demonstrated a comparable decline during a covert assessment condition, despite the fact that the observers were more experienced and fewer target behaviors were being monitored. Romanczyk et al. (1973) also found that the level of observer agreement obtained during overt reliability checks depended upon whether the observer was informed as to who was performing the assessment. The reliability of observers with an identified assessor was consistently higher than reliability with an unidentified assessor.

The first implication from the above studies is that observers do not necessarily remain highly accurate and consistent over time. Although investigators have repeatedly used single behavioral observers, limiting reliability checks
to initial training sessions (e.g., Beckwith, 1972; Osofsky & O'Connell, 1972; Patterson & Reid, 1970) or to some point(s) during actual experimental monitoring (e.g., Berk, 1971; Walker & Buckley, 1968), there is no assurance that observers will remain highly accurate during unassessed periods (Taplin & Reid, 1973). If two independent observers cannot be employed simultaneously, in order to gain continuous reliability feedback, it has been suggested that measures of single observer accuracy (i.e., spot checks) be obtained covertly to prevent spurious reliability inflations (Johnson & Bolstad, 1973; Taplin & Reid, 1973). A second implication (relating to the findings of Romanczyk et al., 1973) is that pairs of observers can modify or drift in their definitions of target behaviors in order to obtain high reliabilities with each other but low reliabilities with another observer using standard behavioral definitions. Although simultaneous observation by two or more observers is highly preferred to the single observer method, caution must be exercised in order to prevent discussion of observational data and the development of idiosyncratic behavioral definitions. Bornstein & Quevillon (1976) have suggested a strategy to deal with observer drift which approximates a constant criterion or "pure calibrator" assessment (Johnson & Bolstad, 1973). In their study, ten covert reliability checks were made by the senior author in order to check for drift away from standard behavioral definitions. Through
the use of such criterion approximations, researchers can either become more confident that drift has not occurred or can retrain or recalibrate observers in order to prevent the continued use of idiosyncratic behavioral definitions.

A third problem facing investigators who employ behavioral observation as a primary assessment tool is that of observer bias (Rosenthal, 1963, 1968). Although a few cases of intentional and unbridled data fabrication have appeared in the literature (e.g., Azrin, Holz, Ulrich, & Goldiamond, 1961), most of the findings relating to observer bias have come from controlled investigations where knowledge of experimental hypotheses has been the manipulated independent variable. One notable exception is the unusual and highly atypical acknowledgment made by Scott, Burton, and Yarrow (1967). In their investigation, it was found that the senior author's observational data differed significantly from that of other blind observers and showed stronger support for the experimental hypothesis under study. Although this uncontrolled finding is subject to criticism on numerous methodological grounds (see Johnson & Bolstad, 1973), other investigators have found similar incidences of observer bias in more tightly controlled studies. For example, Kass and O'Leary (Note 5) reported differences between informed and uninformed observers even though ratings were made from the same set of video tapes. Observers who were in-
formed that the level of disruptive behavior was expected
to decrease showed a biased decline in the amount of dis­
ruptive behavior observed.

Other investigators, however, have failed to find
significant experimental effects attributable to observer
knowledge of expected results. Skindrud (Note 6) divided
observers into three groups and provided differential ex­
pectations regarding the target behaviors of videotaped
family members (i.e., target behaviors would increase, de­
crease, or stay the same). No significant differences were
found in the data obtained from the three observer groups.
In a similar study (Kent, O'Leary, Diament, & Dietz, 1974),
differential expectations (i.e., decrease, stay the same)
produced a nonsignificant effect on the observational data
reported but did exert a significant effect on the observers'
overall "subjective" report of behavior change. While both
of these studies call into question the importance of ob­
server bias as a significant methodological problem, O'Leary,
Kent, and Kanowitz (1975) have subsequently demonstrated
that a combination of observer knowledge of expected results
and verbal feedback regarding the degree to which collected
data fulfill experimental predictions can exert a signifi­
cant biasing effect on the observational reports produced.
Since experimenters or other individuals in the therapeutic
environment may on occasion offer evaluative comments re­
garding observer's recorded data, subsequent observations
may be biased as a result of prior feedback.

Although observer bias as a consistent and powerful methodological problem has not received overwhelming experimental confirmation, the potential for bias resulting from observer knowledge of predicted results does warrant further attention and continued effort to control for its confounding influence. Since high levels of reliability between observers does not theoretically control for observer bias (see discussion by O'Leary et al., 1975), investigators should continue to keep behavioral raters blind to the experimental hypotheses under investigation (Jeffrey, 1974a) and refrain from discussing observational data in their presence (O'Leary et al., 1975). However, further precautions may be necessary in view of the fact that even experimentally blind observers may be able to discern experimental predictions by simply noting the environmental manipulations taking place in the treatment setting (e.g., dispensing of tokens, contingent verbal praise, withdrawal of reinforcement, etc.). In situations where therapeutic manipulations are patently obvious, counter-demand instructions and visible placebo manipulations may prove useful in masking the experimental predictions under investigation (see Hamilton, Quevillon, & Bornstein, in press; Steinmark & Borkovec, 1974).

In conclusion, several significant problems exist for the applied researcher utilizing behavioral observation as
a primary assessment strategy. In addition to the procedural complications created by reactivity, instrument decay, observer drift, and observer bias, a number of additional factors related to observer accuracy have recently been identified. These factors include the complexity of the behavioral categories employed (Mash & McElwee, 1974), the frequency of the behaviors emitted (Patterson & Harris, Note 1), the complexity of the behaviors observed (Reid & Jones, 1974), and the predictability of the observed behaviors (Mash & McElwee, 1974). Although research in the area of behavioral observation continues to point out the complexity and intricacies of the observational process, procedures to circumvent many of the more commonly encountered problems are available and should be employed wherever possible.

Self-monitoring

Self-control or self-management treatment procedures have become increasingly popular during the past five years and have been used to modify a wide range of problem behaviors (see Mahoney & Thoresen, 1974; Thoresen & Mahoney, 1974). Although a large number of specific self-control techniques currently exist (e.g., self-reward, aversive self-regulation, cognitive mediation, etc.), each is grounded on the unitary premise that individuals can be taught to modify their extratherapeutic environment and
to self-administer specific treatment techniques in order to affect change in their own behavior (Kanfer & Phillips, 1970; Kazdin, 1974). In addition to the expansion of the client's role as a behavioral change-agent, self-control procedures typically rely on the client to gather his/her own behavioral data in order to allow for constant feedback regarding the efficacy of self-imposed treatment strategies. Indeed, self-monitoring of problem behaviors has been viewed as an essential component in any self-control program (Bandura, 1971; Buckley, 1968; Glynn & Thomas, 1974; Kanfer, 1971). Not only is self-monitoring highly consonant with a self-control model of treatment, it also offers distinct advantages over external monitoring in that (a) it allows for immediate and continuous behavioral feedback, (b) it allows access to data which is not available to an external observer (e.g., covert behaviors), and (c) it is extremely portable and economical (Kazdin, 1974; Mahoney & Thoresen, 1974). Despite these advantages, serious problems exist for the applied researcher interested in employing self-monitored data as a sole measure of treatment outcome. As Franks and Wilson (1975) have recently stated, "If observer monitoring is suspect, self-monitoring—a strategy often advocated on attractive grounds of economy, convenience, and the possible reduction of some of the methodological problems (in external monitoring)—presents seemingly insuperable difficulties" (p. 244). When used as
an assessment device, self-monitoring is plagued by problems in two major areas; reactivity and reliability.

As noted in the previous section, the obtrusive presence of an external observer often acts as a reactive stimulus which precipitates change in the behavior of the individual being observed (e.g., Patterson & Harris, Note 1; Johnson & Lobitz, Note 2). A similar situation arises in self-monitoring in that clients serve as obtrusive observers of their own behavior. As might therefore be expected, self-monitoring has often been found to bring about change in self-observed target behaviors. Self-monitoring has been shown to evoke reductions in smoking rate (Rutner, Note 7), disruptive behaviors (Broden, Hall, & Mitts, 1971), reported hallucinations (Rutner & Bugle, 1969) and maladaptive motor behaviors (Maletzky, 1974); and has been shown to increase study behavior (Johnson & White, 1971) and the amount of attention paid to appropriate child behaviors (Herbert & Baer, 1972). Despite the fact that self-monitoring often precipitates substantial behavior change, the reactive effects of self-monitoring have not always been consistent and several well controlled investigations have failed to find significant effects attributable to self-observation (e.g., Baer, 1972; Hall, 1972; Jackson, 1972; Mahoney, 1971; Mahoney, Moore, Wade, & Moura, 1973; Powell & Azrin, 1968; Stollak, 1967). Although several theoretical rationales have been offered to explain the reactive effects often noted
(e.g., Franks & Wilson, 1975; Kanfer, 1970), recent research has shown that reactivity of self-monitoring is influenced by the valence (i.e., social desirability) of the behaviors being monitored, whether monitoring occurs before or after the terminal behavior, and by providing performance goals and feedback (Kazdin, 1975b; Romanczyk, Tracey, Wilson, & Thorpe, 1973).

While self-monitoring can often serve in and of itself as a useful treatment procedure, reactivity creates a major problem for researchers interested in employing self-monitoring as a data collection device; namely, multiple treatment interference (Campbell & Stanley, 1966). Since self-monitoring itself can potentially change the behavior (i.e., dependent measure) under investigation, it cannot be used to assess the separate effects of another form of treatment. To do so would result in confounding the effects attributable to both self-monitoring and the treatment procedure employed. To circumvent this problem, two methodological strategies have been suggested (Jeffrey, 1974a; Nelson & McReynolds, 1971; Thoresen & Mahoney, 1974).

In multiple group designs, treatment effects attributable to self-monitoring can be parcelled out by including a no-treatment control group, a self-monitoring only group, and a self-monitoring plus treatment group. In single-subject designs, researchers may either delay treatment until the self-monitored baseline has become sufficiently stable or
may incorporate the following experimental phases into their treatment plan: (a) externally-monitored baseline, (b) self-monitoring, and (c) self-monitoring plus treatment.

While the reactive effects of self-monitoring can be isolated by employing one of the experimental designs mentioned above, the problem of self-report accuracy has presented behavioral researchers with a more formidable methodological task (Kanfer, 1970; Nelson & McReynolds, 1971; Simkins, 1971a, 1971b). Although reliable self-monitoring has not been considered to be a prerequisite for behavioral improvement (Broden et al., 1971; Herbert & Baer, 1972), when used as an assessment procedure, self-monitoring clearly demands accurate and unbiased self-reporting (Jeffrey, 1974a). Unfortunately, several investigations comparing levels of agreement between self- and externally-monitored data tend to show that subjects are rather unreliable self-observers (e.g., Broden et al., 1971; Fixsen et al., 1972; Herbert & Baer, 1972; Hendricks, Thoresen, & Hubbard, Note 8; Lipinski & Nelson, 1974b; McFall, 1970; Thomas, Abrams, & Johnson, 1971; Thoresen, Hannum, Hendricks, & Shapiro, Note 9). This situation has led some researchers to conclude that, "the naive assumption that highly motivated subjects will be both consistent and accurate in their self-reporting is not supported by the available evidence" (Thoresen & Mahoney, 1974, p. 35).

Inaccurate or unreliable self-monitoring has been
attributed to (a) differences in behavioral definitions (i.e., criteria) used by self and independent observers, (b) emergence of prepotent behaviors which interfere or are incompatible with self-recording, and (c) fatigue resulting from continuous self-monitoring of high frequency behaviors (Edelstein & Noah, Note 11; Jeffrey, 1974a; Simkins, 1971a). It has also been repeatedly suggested that demand characteristics (Mahoney, 1974; Thoresen & Mahoney, 1974), evaluation apprehension (Jeffrey, 1974a), or direct reinforcement of hypothesis confirming data (Simkins, 1971a, 1971b; Kazdin, 1974) may serve to bias subjects' self-report in a treatment validating direction. Indeed, therapists may unintentionally alter self-monitored data by providing subtle cues which indicate to the client what degree of behavior change is expected and desired. In order to please the therapist (Orne, 1969) or to project a favorable image (Rosenberg, 1969), the subject may be placed in the precarious position of either intentionally or unintentionally producing acceptable and therapist reinforcing data. Furthermore, even when attempts are made to control for the demand characteristics inherent in the treatment setting, subtle suggestions for improvement may be no less obvious (Kazdin, 1974). As Mahoney (1974) recently stated, "To prescribe a particular self-management technique for an individual is tantamount to asking for a given behavior change" (p. 262). In accordance with the above statements, prior research has demon-
strated that subjects tend to underestimate the performance of undesirable behaviors (Bolstad & Johnson, 1971; Thomas et al., 1971) while overestimating the occurrence of desirable behaviors (Fixsen et al., 1972; Risley & Hart, 1968; Santogrossi, O'Leary, Romanczyk, & Kaufman, 1973). While further research into the potential causes of inaccurate self-monitoring is clearly warranted, much of the current research attention has been directed toward ways to increase the reliability of self-reported data.

Essentially, four broad approaches to the problem of self-monitoring accuracy have been investigated. First, primary reinforcers such as food snacks (Risley & Hart, 1968) or monetary rewards (Lipinski, Black, Nelson, & Ciminero, 1975) have been used to reinforce agreement between self-report and observational records of performance. For example, Lipinski et al. (1975) paid self-observers $1.00 per session for obtained reliabilities ≥ .90 in an attempt to increase the self-monitoring accuracy of face-touching behavior. Subjects unable to meet this criterion were given monetary incentives of four cents per 1 percent increase over baseline accuracy. All subjects were additionally provided with reliability feedback as rewards were being dispensed. The results of this investigation revealed that the average reliability (i.e., agreement between unobtrusive observers and self-recorders) increased from .46 at baseline to .81 following experimental manipulations. While self-monitoring
accuracy was significantly enhanced by reinforcing high levels of reliability or successive approximations to criterion, the reinforcement for accuracy approach would appear problematic for the following reasons: (a) In order to reinforce accuracy of self-monitored data, one must first know when the client is, in fact, being accurate. Unfortunately, this is not always possible, especially in those situations where the target behavior is covert (Bucher & Fabricatore, 1970; Hamilton & Bornstein, Note 10; McFall, 1970; Rutner & Bugle, 1969), where independent observation by external observers is simply unrealistic (Kanfer, 1970), or where no overt concomitants of the private event exist (Kazdin, 1974). (b) In order to demonstrate increased reliability of self-report under natural (i.e., unobtrusive) conditions, subjects must perceive that they are being given the opportunity to independently record behaviors following a reinforcement-for-accuracy condition. Risley and Hart (1968) provided no such condition, while Lipinski et al. (1975) interspersed reinforcement and no-reinforcement conditions during self-observational sessions. These brief independent self-monitoring periods (i.e., subjects were unobtrusively observed but offered no monetary incentives) are clearly inadequate checks on reliability maintenance. In spite of the brevity of the independent monitoring periods, Lipinski et al. showed a 14 percent decline in self-monitoring accuracy as monetary incentives were
periodically withdrawn. (c) In the Lipinski et al. investigation, a decline in the frequency of face-touching covaried with an increase in self-monitoring accuracy. Since it has been previously noted that reliability increases as the frequency of the observed behaviors decreases (e.g., Patterson & Harris, Note 1), it is difficult to maintain that reinforcement and not the reactive effects of self-monitoring produced the increase in accuracy. (d) Target behaviors such as verbalizations of block-building (Risley & Hart, 1968) and frequency of face-touching (Lipinski et al., 1975) would appear to be somewhat lacking in clinical relevance and, quite simply, may not be analogous to the therapeutic situation requiring accurate self-report (i.e., in terms of treatment demand, target behaviors, etc.).

A second but somewhat similar approach to the problem of increasing the reliability of self-monitoring has involved the training of subjects to "match" the report given by external observers by making reinforcement for primary target behaviors (e.g., appropriate social and academic behavior) contingent upon accurate self-monitoring (Bolstad & Johnson, 1972; Drabman, Spitalnik, & O'Leary, 1973; Fixsen et al., 1972; Turkewitz, O'Leary, & Ironsmith, 1975). For example, in the Turkewitz et al. (1975) investigation, a self-reinforcement program was implemented to increase the appropriate social and academic behaviors of school children. In order to improve the accuracy of self-reported data (upon
which the amount of reinforcement was based), a matching and fading procedure was introduced. Initially, all students received the amount of self-determined points (exchangeable for tangible rewards) if they were within one point of the teacher's ratings. If teacher and student ratings matched exactly, the child was given one additional point. However, if there was more than a one point discrepancy, no points were received for that rating period. Each student additionally received accuracy feedback and verbal praise for high reliability. Following the experimental phase in which all children were required to match, three successive fading periods were implemented. During the first fading period, only one-half of the children were required to match; during the second, only one-third of the children matched; and during the third period, all students were given the total amount of self-determined points without matching. The teacher continued to administer accuracy feedback and verbal praise for accurate data throughout all experimental periods. Results revealed the following percent of students within one point of a perfect match across the four experimental phases: (1) total matching = 84%, (2) 1/2 matching = 81%, (3) 1/3 matching = 81%, (4) no matching = 61%. Although the matching procedure was quite effective in increasing accuracy over the level noted at baseline (i.e., 14 percent), it is readily apparent that a decline in reliability takes place as the matching and fading process continues. This
decline in accuracy has been noted in other research as well (i.e., Drabman et al., 1973; Fixsen et al., 1972).

Two additional problems with the matching procedure include: (a) In cases where non-reinforcement related treatment strategies are externally applied (e.g., soft reprimands, systematic desensitization, etc.) or in situations where such contingencies are self-applied (e.g., covert reinforcement, thought stopping, etc.), the matching procedure unfortunately appears inapplicable. (b) Situational practicalities may arise in which external observation for the purpose of matching is simply impossible (e.g., observation of covert events, monitoring behaviors in multiple settings, etc.).

A third approach has utilized stimulus cues to increase the veridicality of self-monitoring. Edelstein and Noah (Note 11) employed sixteen college students to assess the effects of three "cue" conditions on the accuracy of self-monitoring. Three groups of students (monitoring frequency of face-touching) were exposed to one of three orders of the following cue conditions: (1) an independent obtrusive observer, (2) an independent obtrusive observer who was clearly not recording face-touching, and (3) a simple visual cue (i.e., a sign with the word "Record," placed on the blackboard). Results revealed that none of the three cue conditions significantly differed in terms of self-monitoring accuracy. Since previous research has shown that self-observers are significantly more accurate when aware that
reliability measures are being taken (Lipinski & Nelson, 1974b; Nelson, Lipinski, & Black, 1975), these results are quite impressive in that no differences were found between the visual cue condition and a condition in which an obtrusive observer was present. Simple visual cues reminding subject to "record" could prove to be an expedient method of promoting veridical self-monitoring. The main problem, however, is that subjects in the Edelstein & Noah (Note 11) study were highly inaccurate across all conditions, despite the fact that no differences between cue conditions were found. The approximate percent observer agreement score (i.e., between self and unobtrusive external observers) for the visual cue condition was 40 percent. Inter-rater agreement scores at this level are clearly inadequate and allow for ambiguous interpretation of behavior change (Kazdin, 1975a).

A final approach to increasing the fidelity of self-monitored data has been the Reliability Enhancement Package of Bornstein, Hamilton, Carmody, Rychtarik, and Veraldi (Note 12). The Reliability Enhancement Package (REP) is composed of four procedural manipulations (i.e., cognitive consistency, consequence clarification, public commitment, and cueing) which are designed to increase the probability of obtaining accurate self-monitored data. Each procedural manipulation is theoretically grounded and backed by research in the areas of social psychology, cognitive processes,
and learning (e.g., Aronson & Mettee, 1968; Blackwood, 1972; Cook & Insko, 1968; Deutsch & Gerard, 1965; Graf, 1971).

In their study (Bornstein et al., Note 12), 18 subjects, self-referred for a relaxation training program, were assigned to either a Reliability Enhancement Package (REP) group or to an Attention-Control (AC) condition. Subjects in both groups were given a "bogus" biofeedback relaxation assessment before and after four nightly sessions of therapist-administered progressive relaxation training. During the pre- and post-treatment biofeedback assessment, subjects were asked to self-monitor "states of relaxation" by depressing a hand-held toggle switch each time an "alpha burst" was heard over the headphones. The hand-held toggle switch activated a digital timer which made it possible for subjects to self-monitor their time in alpha during each of the 25 relaxation trials administered. The "alpha bursts" were actually standardized pre-recorded tones fed into the headphones from an adjacent room. The same set of tones were used during both pre- and post-treatment assessment sessions.

Subjects in the REP condition received the following four procedural manipulations immediately after the final (i.e., fourth) session of progressive relaxation training but prior to the post-treatment biofeedback assessment:

(1) Cognitive Consistency - REP subjects were given "bogus"
personality feedback from a questionnaire completed prior to the initiation of treatment. A series of Barnum-type personality attributes were presented (Forer, 1949), concluding with the "honesty" attribution: "You are honest in evaluating your own performance and have a high degree of personal integrity." (2) Contingency Clarification - After the personality feedback had been administered, a series of statements clarifying the negative consequences of inaccurate self-monitoring were given. (3) Public Commitment - Subjects received a phone call from an experimental stooge who attempted to elicit information from REP subjects regarding the cognitive consistency and contingency clarification components previously mentioned. (4) Cueing Statement - During the post-treatment biofeedback assessment, REP subjects had the following statement typed at the top of their data sheets: "Remember, it is extremely important to be accurate in recording your scores!"

Subjects in the AC condition were exposed to similar manipulations (i.e., Barnum-type personality attributes, information eliciting phone calls, cueing statement), but without the procedural components designed to increase the accuracy of self-monitoring (i.e., honesty attribution, "accuracy" cue, etc.).

The results of this investigation indicated that REP subjects were significantly more accurate than AC subjects (i.e., in self-monitoring the duration of alpha tones).
during the post-treatment biofeedback assessment ($p < .007$). The mean inaccuracy scores from pre- to post-treatment indicated no change for the REP group, while subjects in the AC condition became significantly more inaccurate ($p < .01$). Furthermore, when deviations occurred at post-treatment assessment for AC subjects, they were in the direction of overestimating time in alpha (i.e., falsely reporting increases in relaxation ability).

These results were interpreted as indicating that AC subjects were responding to the "experimental demand for improvement" inherent in the treatment setting. Since AC subjects increased their total seconds of inaccuracy nearly threefold (i.e., from pre- to post-treatment) and since the direction of inaccuracy shifted in a treatment validating direction (i.e., showing increased relaxation abilities), it may be inferred that the demand for improvement during post-treatment assessment was quite strong and quite pervasive. The four reliability enhancement manipulations apparently minimized post-treatment demand, allowing REP subjects to maintain the level of accuracy obtained during pre-treatment assessment. Rather than attempting to please the therapist (Orne, 1969) or striving to project a favorable self-image (Rosenberg, 1969) by producing self-monitored data which would confirm the effectiveness of the treatment procedures employed, REP subjects seemed quite responsive to the overt and persistent suggestion that accurate recording
and not the demonstration of increased relaxation abilities was the important task during post-treatment monitoring.

In the methodologically ideal self-monitoring situation, clearly defined target behavior(s) are monitored not only by the subject but also by some independently verifiable source (Jeffrey, 1974a; Mahoney, 1974). While unobtrusive external monitoring is the preferred method of estimating the accuracy of self-monitored data, this tactic is simply not feasible when the behavior is covert (e.g., hallucinations, urges to smoke, derogatory self-statements) or in situations where self-monitored data is being gathered in multiple settings (Mahoney & Thoresen, 1974). The pragmatic value of the REP procedure lies in its potential to provide an alternative means of facilitating self-report accuracy without first observing and then reinforcing its occurrence (e.g., Lipinski, et al., 1975; Turkewitz et al., 1975). While it is impossible to guarantee that subjects receiving REP manipulations will be consistently accurate and honest in their self-evaluations, the probability of obtaining accurate and reliable self-reports would appear greatly improved. The use of "reliability enhancers" should therefore be encouraged in those situations where corroborative data does not exist; particularly in innovative case studies (Lazarus & Davison, 1971) and/or exploratory self-control research (Jeffrey, 1974a). Such preliminary research may then lead to more tightly controlled investigations as a further test
of experimental hypotheses.

Despite the potential applicability of the Reliability Enhancement Package (Bornstein et al., Note 12), several questions regarding this procedure remain to be addressed:

1. In the Bornstein et al. (Note 12) investigation, accuracy of self-monitoring was assessed by having subjects monitor pre-programmed responses by depressing a hand-held toggle switch each time an "alpha" tone occurred. Although subjects believed that the tones were actually auditory representations of their own alpha waves, the self-monitoring task employed was quite different from typical monitoring tasks in that subjects were able to direct undivided attention to self-recording responsibilities. In a more typical self-monitoring situation, specific target behaviors are designated and the subject is then responsible for monitoring their occurrence in the complex extratherapeutic environment. The subject must not only recognize that a target behavior has been emitted and record its occurrence, but he/she must do so while responding concurrently to other stimuli and situations in the natural environment. While REP was found to be quite effective in maintaining the accuracy of self-monitoring when simple pre-programmed responses were employed, this does not necessarily imply that these results will generalize to the situation where observed behaviors are self-emitted and where the subject must continue to simultaneously respond to non-monitored
stimuli. The question of generalizability becomes even more pertinent in view of recent research which suggests that accuracy of self-monitoring decreases as concurrent operant tasks are introduced (Epstein, Webster, & Miller, 1975; Epstein, Miller, & Webster, 1976). Since it has been proposed that REP operates by decreasing treatment demand and thus allowing subjects to maintain initial accuracy levels, it would appear highly probable that in more complex monitoring situations, the initial level of accuracy would be quite low (e.g., Broden et al., 1971; Lipinski & Nelson, 1974b; McFall, 1970; Thoresen et al., Note 9). Although REP may prevent the reliability of self-monitored data from declining even more as a result of bias created by post-treatment demand, subjects may continue to be highly unreliable despite the fact that demand has been minimized. In these situations, alternate forms of self-monitoring training may be necessary to increase initial accuracy levels before reliability enhancement procedures are implemented.

2. While REP was found to be significantly more effective in promoting accurate self-evaluations than Attention Control (AC) manipulations, the question remains as to whether the same effects could be obtained through less elaborate means. It could be potentially argued that identical effects would be produced by simply telling subjects to be accurate and by providing less intricate record-
ing cues (e.g., the "Record" cue used by Edelstein & Noah, Note 11). If less elaborate procedures produce similar levels of self-report accuracy, applied researchers would naturally want to employ the more parsimonious of the two reliability strategies.

The Present Investigation

The intent of the present investigation was to (a) assess the effectiveness of the Reliability Enhancement Package (composed of cognitive consistency, consequence clarification, and cueing components) in a more typical and complex self-monitoring situation, (b) compare REP to a more elementary procedure composed of accuracy instructions and a simple "Record" cue, (c) assess the combined (and separate) effects of REP plus a training procedure designed to increase the initial level of self-monitoring accuracy, and (d) evaluate the effectiveness of experimental manipulations when the demand for improvement is either "low" (unmanipulated) or "high" (manipulated).

Four criteria were considered in selecting target (i.e., self-monitored) behaviors for the current study. It was considered essential that the behaviors (1) be clinically relevant, (2) be self-emitted rather than pre-programmed, (3) occur concurrently with other non-monitored environmental stimuli, and (4) be amenable to unobtrusive external observation. In view of their ability to meet these criteria,
two categories of normal speech disfluencies were selected as target behaviors, (a) "ah" sounds, and (b) verbal repetitions. Although more than eight different disfluency categories have been previously identified (see Mahl, 1956), research has indicated that "ah" sounds (i.e., verbalized pauses) and repetitions (i.e., superfluous duplications of syllables, words, and phrases) tend to occur more frequently than many of the other disfluency categories (Baker, 1964). Not only has research shown that audience ratings (i.e., ratings of speaker confidence, credibility, and dynamism) drop as the frequency of these two verbal behaviors increase (McCroskey & Mehrley, 1969; Miller & Hewgill, 1964; Sereno & Hawkins, 1967), speech disfluencies have been considered important clinically because of their topographical similarity to moments of stuttering (see Bloodstein, Alper, & Zisk, 1965; Goldiamond, 1965; Siegel, Lenske, & Broen, 1969). Normal speech disfluencies have been successfully treated through the use of a variety of response contingent stimuli, including shock, loud noises, delayed auditory feedback, door buzzers, and the word "wrong" (e.g., Goldiamond, 1965; Martin, 1968; Siegel & Hanson, 1972; Siegel & Martin, 1965, 1966, 1968). Additionally, Siegel (1973) has demonstrated that subjects can suppress speech disfluencies after merely being told that specific disfluent behaviors are being externally monitored.

Subjects for the current investigation were selected
from a population of speech anxious undergraduates who had indicated an interest in participating in a speech anxiety treatment program. While reliability of self-monitoring rather than actual treatment effects was the variable of empirical interest, a clinically relevant analogue (i.e., speech anxious subjects participating in a treatment program) was constructed so that the obtained results would be potentially generalizable to the treatment situation. Subjects were informed that treatment would consist of repeated exposure (i.e., flooding) to the actual feared situation (i.e., public speaking) and that they would be responsible for evaluating their progress by self-monitoring objective signs of anxiety (i.e., frequency of "ah's" and repetitions) as well as subjective perceptions of fear (i.e., internal sensations of anxiety along an 11-point scale) during each speech performance. Subjects self-monitored objective and subjective behaviors during three separate speech sessions (SM I, II, and III). The three speeches given during each speech session were unobtrusively recorded, which allowed for a comparison to be made between the number of disfluencies self-monitored and the number actually emitted. Experimental manipulations were administered between SM I and SM II. Each subject received one set of instructions: Reliability Enhancement Package (REP) or Accuracy Instructions (AI), and one set of training procedures: Self-monitoring Training (SMT) or Training Control (TC). Experi-
mental demand for improvement was manipulated prior to the final speech session (SM III). All subjects were essentially told that they could expect large declines in anxiety (objective and subjective) during the final set of three speeches.

It was hypothesized that (1) a significant decline in the number of self-monitoring errors would occur between SM I and SM II for subjects receiving Self-monitoring Training (SMT), (2) a significant decline in the number of self-monitoring errors would not occur between SM I and SM II for subjects receiving the Training Control (TC) procedure, (3) Accuracy Instructions (AI) and the Reliability Enhancement Package (REP) would not significantly affect the number of errors made during SM II, and (4) when treatment demand was manipulated during SM III, REP would maintain the accuracy level obtained during SM II, whereas AI subjects would make significantly more self-monitoring errors. Experimental predictions are graphically presented in figure 1.
Figure 1. Approximate number of self-monitoring errors predicted for the four experimental conditions across the three self-monitoring periods.
CHAPTER II

METHOD

Subjects

Subjects for the current investigation were selected from 360 University of Montana introductory and developmental psychology students who had previously completed Paul's (1966) short form of the "Personal Report of Confidence as a Speaker" (PRCS, see appendix A). Students with the highest speech anxiety scores (i.e., PRCS range = 28 - 21) were contacted and asked to participate in a treatment program for public speaking anxiety. Prospective subjects were further informed that the treatment program (offered through the psychology department) would focus on the subjective and objective manifestations of speech anxiety and that research credit would be allowed for participation. Of the 52 students contacted, 36 subjects (13 men, 23 women) indicated an interest in participating and were able to attend treatment sessions as scheduled. Each subject was randomly assigned to one of 12 experimental groups, with three subjects serving within each group. All subjects were requested to complete the "Eysenck Personality Inventory" (Eysenck & Eysenck, 1963) before attending the initial treatment session.
Experimental Design

The basic design of the present experiment is depicted in Table 1. The twelve experimental groups represented a 2x2x3x3 factorial design for repeated measures with training (Self-monitoring Training vs. Training Control), instructions (Reliability Enhancement Package vs. Accuracy Instructions), and therapists (one vs. two vs. three) serving as

<table>
<thead>
<tr>
<th>Experimental Conditionsa</th>
<th>Therapists</th>
<th>N</th>
<th>SM I</th>
<th>SM II</th>
<th>SM III</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMT-REP</td>
<td>T₁</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₂</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₃</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-AI</td>
<td>T₁</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₂</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₃</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC-REP</td>
<td>T₁</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₂</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₃</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC-AI</td>
<td>T₁</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₂</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₃</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aExperimental conditions were manipulated during the second experimental phase (i.e., Instruction and Training) which occurred between SM I and SM II.
between-subjects variables and self-monitoring periods (Self-monitoring I vs. Self-monitoring II vs. Self-monitoring III) as a within-subjects factor. More specifically, nine of the 36 total subjects were assigned to one of the following four factorial combinations: (1) Self-monitoring Training (SMT) and Reliability Enhancement Package (REP), (2) Self-monitoring Training (SMT) and Accuracy Instructions (AI), (3) Training Control (TC) and Reliability Enhancement Package (REP), and (4) Training Control (TC) and Accuracy Instructions (AI). Three subjects within each of the above combinations were assigned to one of three therapists (T₁, T₂, T₃). All 36 subjects participated in the three self-monitoring phases (SM I, SM II, and SM III).

Therapists

Three advanced (i.e., third year) male graduate students in clinical psychology served as therapists. Each therapist received specific training in administering each of the four experimental conditions (i.e., SMT, TC, REP, and AI) and in following general experimental procedures (i.e., administering the treatment rationale, demand manipulation, etc.). Training involved the use of procedural outlines and verbatim transcripts as well as author directed role-playing and coaching. Mean training time for each therapist was five hours.
EXPERIMENTAL PHASES

Group sessions were held during late afternoons and evenings at the University's clinical psychology center. Each of the twelve groups (consisting of one therapist and three subjects) met for a total of four sessions corresponding to the four experimental periods. During three of these periods (i.e., SM I, II, and III) the procedure was identical for each of the twelve groups. However, during the second period (i.e., Instruction and Training) four experimental conditions were manipulated (i.e., SMT-REP, SMT-AI, TC-REP, TC-AI). Each therapist administered all four of these experimental conditions; one to each of his four groups. The average time delay between Self-monitoring I and the following three periods (i.e., Instruction and training, SM II; and SM III) was 3, 7, and 11 days respectively.

Self-Monitoring I (Session 1)

During the first 45 minutes of the initial session, subjects received an avoidance conditioning rationale for the maintenance of public speaking anxiety (Bandura, 1969) and a brief theoretical explanation (i.e., classical extinction) of in vivo flooding (Malleson, 1959). More specifically, subjects were told that a conditioned emotional reaction (such as public speaking anxiety) can be weakened and eventually eliminated through repeated
exposure to the actual feared situation without "reinforcement" or adverse consequences (D'Zurilla, Wilson, & Nelson, 1973). It was further explained that in vivo flooding procedure to be employed during the current program would involve exposing participants repeatedly to "real-life" public speaking situations. Subjects were told that during the latter part of the present session and during two subsequent sessions (i.e., SM II and III) each of them would be taken to separate clinic rooms where they would deliver three 2 1/2 minute speeches on various preselected topics in front of a two-person audience.

Subjects were then told that an on-going evaluation of the present treatment program would be implemented. It was explained that because of the time and expense involved in employing trained observers to rate the speech behaviors of treatment participants, each subject would be responsible for monitoring his/her own behavior during each speech. More specifically, subjects were informed that two classes of anxiety responses would be self-monitored: (a) objective responses (i.e., verbal disfluencies within two categories; "Ah" and Repetition) and (b) subjective responses (i.e., internal sensations of anxiety along an 11-point scale). Subjects were then provided with a sheet defining each of the two disfluency categories and both were discussed in detail. "Ah" and Repetition disfluencies were defined as follows:
1. "Ah". This category represents the verbalized pause or interjection disfluency. An "Ah" is indicated by the utterance of the "ah", "er", "um", "hu" or "hummm" sound between two words, two syllables, or a syllable and a word. For example, each of the following would be scored as one "Ah" disfluency (Repetitions not included):

"...because of the fact that 'uh' Nixon resigned."
"I'm not really sure if any- 'um' any- anything can..."
"...it would be some-'er' something that..."

If two or more "Ah" verbalizations occur sequentially, each would be scored separately. Sequential "Ah" verbalizations are not tabulated as Repetitions (see the following definition of a Repetition). The following would be scored as two "Ah" disfluencies:

"I'm really not 'uh' 'um' angry, I just feel..."

2. Repetition. A repetition is defined as the serial superfluous repetition of a syllable, word, or phrase. For example, each of the following would be scored as one Repetition:

"It seems to me that we must re- realize that..."
"One doesn't have---have to be an idealist to recognize..."
"It would appear that something---that something needs to be done."

An "Ah" disfluency does not nullify a Repetition. The following statement would be scored as one "Ah" and one Repetition:
"Stricter enforcement of 'uh' of the ruling is necessary."

Multiple Repetitions are scored accordingly. The following would be an example of two Repetitions and one "Ah" disfluency:

"I'm not really sure if anything can be done to help."

In the case of a phrase repetition involving an "Ah" disfluency, the Repetition and "Ah's" would be tabulated separately. The following would be scored as one Repetition and two "Ah" disfluencies:

"'Uh' the Alaska pipeline...'uh'...the Alaska pipeline has both positive and negative aspects."

A word repeated for emphasis is not scored as a Repetition. For example:

"The mountains were very, very beautiful!"

After both disfluency categories had been explained and defined, subjects were presented with copies of the three data sheets which were to be used for self-recording purposes. Data sheet one (see appendix B) contained a set of disfluency category headings (i.e., "Ah" and Repetition) each followed by successive bracketed numbers (i.e., [1] to [24]); one set was provided for each of the three speeches. Subjects were instructed to place a pencil mark over successive numbers, in the appropriate disfluency category, each time an "Ah" or Repetition occurred. Data sheet two (see appendix C) consisted of three 11-point Likert-type scales on which sub-
jects were to rate the amount of subjective anxiety they had experienced during each of the three speeches (i.e., 0 = no anxiety whatsoever, 10 = panic). Subjects were to make their ratings immediately after each speech had been completed. Data sheet three (see appendix D) was essentially a data summary sheet on which subjects were to record the number of Repetitions, "Ah's", and anxiety increments indicated on data sheets one and two. Provision was made on this sheet for a summary of the self-monitored data from each of the three speech sessions (i.e., SM I, II, and III). This allowed each participant to gain continual feedback on his/her progress throughout the program. Subjects were instructed to complete the appropriate section of the summary sheet (i.e., under session I) after all three of the present speeches had been completed and to hand it to their respective therapists before leaving the clinic. Subjects were told to dispose of data sheets one and two. It was emphasized that the data received from self-monitoring was the only means by which therapists would be able to evaluate the effectiveness of the current program and that audience members were simply "warm bodies" (i.e., undergraduates who had been given no information whatsoever concerning the program) who were being used solely to make the public speaking situations more realistic.

When each therapist was convinced that all three of his group members thoroughly understood the self-monitoring
procedure, each participant was given a set of three speech cards which listed each of the designated topics and several subtopics (see appendix E). Subjects were then taken to separate clinic rooms where two undergraduate students (one male, one female), serving as audience members, were seated approximately 2.3m in front of a standard office desk. Situated on top of the desk was a lectern and a cassette tape recorder. After the therapist had positioned the sub­ject behind the lectern and had activated the tape recorder, he left the room. Subjects then received the following progression of taped instructions:

I would like you to speak on the following topic for 2 1/2 minutes: (topic given). Please try to keep talking for the entire period even if you feel you are beginning to run out of things to say. You will have 30 seconds to organize your thoughts before I give the signal to begin. You will also be given a signal to stop when your 2 1/2 minutes are up....(30 second pause) ....Please begin your speech now...(2 1/2 minute pause)....OK, you may now stop speaking. Please indicate your rating on data sheet two...(15 second pause)....For your second 2 1/2 minute speech, I would like you to speak on the follow­ing topic: (second topic given). Once again, please attempt to keep talking for the entire time period. You will have 30 seconds to organize your thoughts...(30 second pause)....Please begin your speech now...(2 1/2 minute pause)....OK, you may now stop speaking. Please indicate your rating for speech two on the second data sheet...(15 second pause)....For your third speech, I would like you to speak on the following topic: (third topic given). Please attempt to keep talking for the entire 2 1/2 minutes. You will have 30 seconds to pull your thoughts together...(30 second pause) ....Please begin your speech now...(2 1/2 minute pause)....You may now stop speaking.
Please indicate your rating for speech three on the second data sheet...(15 second pause)....The two audience members may now leave the room....

The speaker should transfer his/her scores from data sheets one and two to the data summary sheet. Be sure you have entered your summary scores under the correct headings before leaving the speech room. Remember to turn your summary sheet in to your therapist before leaving the clinic; you may dispose of data sheets one and two.

After completing the 15 minute speech session, subjects turned their data summaries in to their respective therapists and left the clinic. Therapists offered no comments regarding the self-monitored data when received.

The assigned speech topics for SM I were as follows: (a) What are your opinions regarding the Women's Liberation movement? (b) What are your opinions regarding sororities and fraternities? (c) Is there life on other planets? (see appendix E for subtopics). Audience members (i.e., three males, three females) were recruited from introductory psychology classes and each received experimental credit for participation. One audience pair (i.e., one male, one female) was randomly assigned to one of the three group members from each of the twelve experimental groups. Audience members were instructed to maintain eye contact with the speaker, to sit up straight in their chairs, to avoid excessive head nodding, and to maintain a relatively "deadpan" facial expression.

In order to unobtrusively record the verbal behavior (i.e., frequency of "Ah's" and Repetitions) of subjects
during their three speeches, a battery-operated cassette recorder and microphone was concealed in each subject's desk top lectern. Entry into the lectern was made through a hinged side panel and the interior was insulated to guard against obtrusive recorder noises. Three suction cups secured to the bottom of each lectern made them immobile and the slanted top provided subjects with a surface on which to self-record (i.e., mark their data sheets). Each microphone was positioned securely in a hole in the top of the lectern. Contact paper placed over the lectern top made the hole unnoticeable.

Instruction and Training (Session 2)

During the second (1 1/2 hour) session, subjects received one of four factorial combinations of experimental conditions (i.e., SMT-REP, SMT-AI, TC-REP, TC-AI). Presentation of instructions (REP or AI) preceded training (SMT or TC) in all cases.

For the first part of the second session, therapists met individually (for 10 minutes) with each of their three group members and presented one set of experimental instructions (REP or AI). Following the presentation of instructions, the group rejoined and one set of training procedures was administered (SMT or TC). Each of the manipulations employed during the instruction and training period is explained more fully below.
Reliability enhancement package (REP). Subjects were taken individually to a clinic room in order to be given feedback from the personality questionnaire (i.e., "Eysenck Personality Inventory") they had completed prior to the initial treatment session. A series of Barnum-type personality attributes (Forer, 1949) were subsequently presented by their respective therapists (e.g., "You have a great need for other people to like and admire you"; "You pride yourself as an independent thinker") concluding with the salient remark, "You are honest in evaluating your own performance and have a high degree of personal integrity."

After hearing these statements about themselves, REP subjects were told that their personality characteristics were of particular importance in obtaining an accurate evaluation of the speech anxiety treatment program. Emphasis was placed on the value of self-monitored data and subjects were told that since "the information you give tends to be extremely accurate and truthful, any modifications in the treatment procedure we may make are apt to be a direct reflection of the data you report." It was stressed that inaccurate reporting of data would result in a waste of time, money, and energy. Subjects were further informed that inaccurate reporting of data could cause the program directors to incorporate inefficient and even detrimental components into future programs, thereby seriously limiting the potential gains to be derived by subsequent treatment.
participants.

Finally, during the second and third self-monitored speech sessions (i.e., SM II and III), the following statement was typed in red at the top of each REP subject's three data sheets: "Remember, it is extremely important to be honest and accurate in monitoring and recording!"

Accuracy instructions (AI). As with subjects in the REP condition, AI subjects were taken individually to a clinic room in order to be given personality questionnaire feedback. Subjects were given the same Barnum-type personality attributes presented to REP subjects with the exception of the target comment, "You are honest in evaluating your own performance and have a high degree of personal integrity." The following statement was substituted for the "honesty" attribution: "You prefer a certain amount of change and variety and become dissatisfied when hemmed in by restrictions and limitations" (Forer, 1949).

Before leaving the room, each AI subject was given the following accuracy instructions: "I wanted to remind you that the data we receive from your self-monitoring is the only record we will have of your performance throughout the current treatment program, so please try to be as accurate as possible in recording your data" (adapted from Jeffrey, 1974b; Taplin & Reid, 1973).

Finally, AI subjects had the following cue (see Edelstein & Noah, Note 11) typed in red at the top of all data
sheets used during the second and third speech sessions (i.e., SM II and SM III): "Record".

**Self-monitoring training (SMT).** After SMT subjects had received either the REP or AI manipulation, each therapist met with his three subjects jointly to implement the self-monitoring training procedure. SMT subjects were told that training was being offered in order to aid participants in becoming more aware (i.e., cognizant) of their verbal behavior. The procedure was essentially divided into two separate training phases: (a) external-monitoring training and (b) self-monitoring training.

During external-monitoring training, subjects were initially given data sheets on which to monitor and were separated from one another in order to insure independent recording. A 3 3/4 minute tape recording of a speech containing both "Ah" and Repetition disfluences was then played. The recorded speech was divided into four segments, each representing a progressively longer time period (i.e., 15 seconds, 30 seconds, 1 minute, 2 minutes). Each segment was played three times before moving to a subsequent speech segment. Subjects monitored "Ah" disfluencies during the first playback, Repetitions during the second, and followed along on a typed transcript (of the particular segment monitored) during the third playback. "Ah" and Repetition disfluencies were clearly indexed on each segment transcript (see appendix F) and the total number of disfluencies
emitted during the speech segment was indicated. Immediately after therapists handed out segment transcripts, subjects were requested to compare their self-monitored scores (for both "Ah's" and Repetitions) with the scores indicated on their transcripts. Subjects were asked to raise their hands if they were within (+ or -) 1 unit of the indicated score and were provided with contingent social reinforcement (i.e., "good," "excellent"). After all four speech segments had been completed, a second 3 3/4 minute tape was played. This tape was also divided into four segments (i.e., 15 seconds, 30 seconds, 1 minute, 2 minutes), although each segment was played only twice. Subjects monitored "Ah's" and Repetitions simultaneously during the first playback and followed along on a segment transcript during the second. The same feedback and reinforcement procedures were employed during the simultaneous monitoring phase. Table 2 provides a summary of the entire external-monitoring training procedure.

After subjects had completed all of the external-monitoring training phases, the self-monitoring training segment was initiated. Subjects were taken to individual clinic rooms and instructed to give one 3-minute speech (on a self-selected topic) into a tape recorder. During the speech, subjects were to self-monitor both "Ah's" and Repetitions simultaneously. After the speech had been completed, subjects were instructed to rewind the tape,
### TABLE 2
SUMMARY OF EXTERNAL-MONITORING TRAINING PROCEDURE

<table>
<thead>
<tr>
<th>External-monitoring Training Phases</th>
<th>Actual Number of Recorded Disfluencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 sec</td>
</tr>
<tr>
<td>One disfluency category(^a)</td>
<td></td>
</tr>
<tr>
<td>&quot;Ah&quot;</td>
<td>3</td>
</tr>
<tr>
<td>Repetition</td>
<td>4</td>
</tr>
<tr>
<td>Two disfluency categories(^b)</td>
<td></td>
</tr>
<tr>
<td>&quot;Ah&quot; - Repetition</td>
<td>2-3</td>
</tr>
</tbody>
</table>

\(^a\)Tape 1 ("Ah" disfluencies were monitored during the first playback and Repetitions during the second.

\(^b\)Tape 2 (Both "Ah" and Repetition disfluencies were monitored simultaneously).

externally monitor what they had already self-monitored, and compare the two scores on both "Ah's" and Repetitions.

**Training control (TC).** After TC subjects had received either the REP or AI manipulation, therapists met with subjects as a group in order to implement the training control procedure. TC subjects were informed that training was being offered in order to aid participants in becoming more aware of their verbal behavior. Subjects were further told that this was to be accomplished by having them attend initially to the verbal behaviors of another speaker and subsequently to their own verbal behaviors in a non-threatening situation.
As in the SMT procedure, TC subjects listened to two 3 3/4 minute tape recordings (identical to those used in SMT) of speeches containing both "Ah" and Repetition disfluencies. Subjects were provided with three 11-point Likert-type scales (see appendix G) on which to rate (a) level of verbal disfluency, (b) level of speaker anxiety, and (c) level of speech organization. Participants were instructed not to count (i.e., externally-monitor) the number of disfluencies emitted but rather to focus their attention on the subjective impact of each speech. The first tape was played three times (nonstop), with subjects rating level of disfluency during the first playback, speaker anxiety during the second, and level of organization during the third. The second tape was played twice, with subjects rating all three dimensions during the first playback and adjusting their ratings on the second. Subjects were not provided with differential feedback regarding their ratings.

Following the completion of the first TC phase, subjects were taken to individual clinic rooms and told to give one, 3-minute speech (on a self-selected topic) aloud to themselves (i.e., a tape recorder was not provided as in SMT). During the 3-minute speech, subjects were to self-monitor both "Ah" and Repetition disfluencies simultaneously.
Self-monitoring II and Demand Manipulation (Session 3)

During the first 15 minutes of session three, subjects were taken to individual clinic rooms where the second set of self-monitored speeches were given. The procedure was exactly the same as that followed during SM I. The assigned speech topics for SM II were as follows: (a) Should all forms of gambling be legalized in Montana? (b) Should all women have a legal right to abortion? (c) Should marijuana be legalized? (see appendix E for subtopics.)

After each subject had completed his/her three speeches and marked the data summary sheet, each therapist met briefly with the entire group in order to deliver the following comments:

Before leaving today, I wanted to provide you with some idea of the gains you can expect during the third and final speech session. As you may already know, in vivo flooding has been employed as a treatment procedure in many anxiety-related treatment programs over the past several years. In most cases, participants such as yourselves have been used to collect data and to provide feedback to therapists regarding the effectiveness of this procedure. Thus far, the results have been overwhelmingly positive in that participants have consistently reported dramatic declines in anxiety as treatment sessions progressed. However, in reviewing this research, before putting the current program together, we noticed that the largest drop in anxiety typically takes place after the second flooding session. Although the psychology literature tends to show that a partial decline occurs between the first and second session, the largest and most significant decline consistently arises during the third treatment session. Even people who appear extremely nervous and anxious during the second "real-life" encounter seem to be able to remain calm and
relaxed during the third such encounter. That is essentially why we settled on three treatment sessions as a limit for the current program. We figured, why go to more than three flooding sessions if people can remain relaxed and verbally fluent during the third public speaking session. Anyhow, I just wanted to provide you with this information so that you would realize why the decision was made to stop treatment after session three. I will be anxious to look at your self-monitored data after the next session to see if our program has been as effective as others. I'm confident that it has.

Self-monitoring III (Session 4)

Subjects completed the final set of self-monitored speeches during the fourth session. The procedure was exactly the same as that followed during SM I and II. The speech topics assigned for SM III were as follows: (a) Should we continue to fund athletic programs at the University of Montana? (b) Should the government appropriate more money for national defense? (c) What are your opinions regarding Gerald Ford's performance as President of the United States? (See appendix E for subtopics.)

After subjects had completed their three speeches and turned their data summaries over to their respective therapists, each participant was administered a post-treatment PRCS (see appendix A), a short post-experimental questionnaire (see appendix H), and was subsequently debriefed.

TAPE RATER AND DEPENDENT MEASURES

One female undergraduate psychology major (naive to the experimental design) served as criterion tape rater
during the present investigation. She received approximately five hours of training in recording the frequency of "Ah" and Repetition disfluencies through the use of an expanded version of the training procedure employed during SMT (see external-monitoring training phase). The rater was required to meet a reliability criterion (i.e., Pearson Product-Moment Correlation) of .90 on four consecutive pre-rated practice speeches before initiating experimental monitoring.

The criterion rater was responsible for rating each of the speech samples taken during the present study. To increase the accuracy of external-monitoring, each of the 2 1/2 minute taped speeches were rated twice, once for frequency of "Ah's" and once for frequency of Repetitions. Covert reliability spot checks (see Johnson & Bolstad, 1973; Taplin & Reid, 1973) were conducted on 43 percent of the speech samples by correlating the frequency count made by the criterion rater with the frequency count made by the present author. During the author-conducted spot checks, precautions were exercised to avoid experimenter bias (i.e., the author was blind to the criterion-rater's scores, the subjects' scores, and to the group identity of the speeches rated). Reliability coefficients were calculated on the ratings of three disfluency measures (i.e., "Ah's", Repetitions, and both categories combined).
The primary measure of self-monitoring accuracy used during the present investigation was the number of monitoring errors committed (i.e., on "Ah's", Repetitions, and both categories combined) during each of the three self-monitoring periods (i.e., SM I, SM II, and SM III). Error scores for each self-monitoring period were arrived at by (a) subtracting the number of disfluencies indicated by the subject from the number indicated by the criterion rater, (b) repeating this process for each of the subject's three speeches, and (c) summing the absolute values of the three deviation scores. This procedure yielded three error scores for each subject for each of the three self-monitoring periods (i.e., one for "Ah's", Repetitions, and both categories combined). Two subsidiary measures of self-monitoring accuracy were also used: (a) percentage of speeches with perfect or one disfluency deviation between the self-recorder and the criterion rater, and (b) mean correlations between the number of self- and externally-monitored speech disfluencies.

In order to check for possible changes in the self-report of subjective (i.e., non-observable) behaviors, analyses were also conducted on pre-post PRCS scores and mean anxiety ratings (i.e., the 11-point Likert ratings) for each of the three experimental phases. Although these ratings were not amenable to external verification, differences in anxiety ratings could potentially be produced by the
different accuracy manipulations employed. Furthermore, in order to check for differences in the "actual" number of disfluencies emitted as a result of procedural manipulations (i.e., REP, AI, SMT, TC), disfluency rates (externally-monitored) were also subjected to statistical analyses.
CHAPTER III

RESULTS

Subject Loss

Self-monitored data from six subjects were lost during the course of the present investigation. Three subjects failed to attend scheduled sessions due to illness, one subject dropped out of school, one returned home because of a death in the family, and one subject was lost due to tape recorder malfunction. In order to maintain an equivalent number of subjects per experimental condition, one subject from both the REP-SMT and REP-TC conditions was randomly eliminated. The results of the present study are therefore based on the data from the 28 remaining subjects (10 males, 18 females).

Inter-rater Reliability

Covert reliability checks were conducted on 43 percent of the ratings (i.e., frequency counts of "Ah" and Repetition disfluencies) made by the criterion rater. More specifically, in order to increase the representativeness of the reliability measures obtained, spot checks were conducted on the taped speeches of three randomly selected subjects from each of the four factorial treatment com-
binations. As previously mentioned, inter-rater reliability was calculated by correlating the frequency counts (on "Ah's", Repetitions, and both disfluency categories combined) made by the criterion rater with those made by the present author. Pearson correlation coefficients were computed separately for each of the four experimental conditions and then averaged via $z$ scores (see Edwards, 1950). Mean correlations between the disfluency ratings made by the criterion rater and those made by the present author were as follows: (a) "Ah" disfluencies = .99, (b) Repetition disfluencies = .97, and (c) both categories combined = .98. It may be concluded that a high degree of consistency in rating existed between the criterion rater and the spot checker.

Change in Number of Self-monitoring Errors

In order to test the equivalency of groups prior to the introduction of experimental manipulations, three separate 2 (SMT vs. TC) x 2 (REP vs. AI) ANOVA's were conducted on pre-manipulation (i.e., SM I) error scores. Results indicated no significant main effects or interactions for errors committed on "Ah's", Repetitions, and both disfluency categories combined (smallest $p > .10$). Pre-manipulation means for the number of self-monitoring errors on "Ah's" and Repetitions combined were as follows: SMT-REP ($M = 19.71$), SMT-AI ($M = 18.00$), TC-REP ($M = 20.00$), TC-AI ($M = 12.14$).
In order to evaluate change in number of self-monitoring errors committed from pre-manipulation (SM I) to post-manipulation periods (SM II and SM III), error scores (for SM I, II, and III) were converted into two separate change scores. The first score (SM II minus SM I) represented change in self-monitoring errors resulting from training and instruction. The second score (SM III minus SM I) represented change in self-monitoring errors which had been maintained after demand for behavior change had been manipulated (see p. 49). However, maintenance of experimental effects over time is intrinsic to both change scores in that SM II occurred four days after training and instruction while SM III occurred eight days after training and instruction.

To simplify subsequent change score analyses, an initial assessment of therapist differences was conducted. The resulting 3 (therapists) x 2 (change score) repeated measures ANOVA for unequal n revealed no significant differences between therapists on change in number of self-monitoring errors committed by their respective group members. These results were consistent for errors on "Ah" disfluencies, Repetitions, and both categories combined (smallest p > .10).

In view of the nonsignificant therapist effect, this factor was eliminated in all subsequent analyses. Two, 2 (SMT vs. TC) x 2 (REP vs. AI) change score ANOVA's were conducted to evaluate differential decline in number of
self-monitoring errors on "Ah's", Repetitions, and both disfluency categories combined. The first analysis was for change between SM I and SM II and the second between SM I and SM III. Change score means for each experimental condition are presented in table 3.

"Ah" disfluencies. A 2 (training) x 2 (instruction) ANOVA for change in self-monitoring errors from SM I to SM II revealed a significant main effect for training \( [F (1, 24) = 5.89, p < .03] \) and instruction \( [F (1, 24) = 4.22, p < .05] \), but a nonsignificant training x instruction interaction \( (F< 1) \). Evaluation of change score means indicated that subjects receiving SMT manipulations declined in number of self-monitoring errors committed (\( M = -5.15 \)) whereas TC subjects increased in number of "Ah" disfluency errors (\( M = 1.36 \)). The data also indicated that REP subjects declined in number of self-monitoring errors (\( M = -4.65 \)) while AI subjects increased in number of "Ah" disfluency errors (\( M = .86 \)). Although group interactions failed to achieve significance, one-tailed correlated \( t \) tests revealed that SMT-REP was the only condition to show a significant within group decline in number of errors from SM I to SM II (see table 3).

A 2 x 2 ANOVA for change in "Ah" disfluency errors from SM I to SM III revealed that SMT (\( M = -3.43 \)) was no longer significantly different from TC (\( M = -.29 \)), [\( F (1, 24) = 1.01, p < .33 \)] although a trend favoring instructional
TABLE 3
CHANGE IN MEAN NUMBER OF SELF-MONITORING ERRORS FROM PRE-MANIPULATION PERIOD (SM I)

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Self-monitoring period</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SM II</td>
<td>M</td>
<td>SD</td>
<td>t</td>
<td>SM III</td>
<td>M</td>
</tr>
<tr>
<td>&quot;Ah's&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>-7.29</td>
<td>5.94</td>
<td>-3.25****</td>
<td>-6.29</td>
<td>6.18</td>
<td>-2.69***</td>
</tr>
<tr>
<td>SMT-AI</td>
<td>-3.00</td>
<td>5.54</td>
<td>-1.44</td>
<td>-0.57</td>
<td>9.02</td>
<td>-0.17</td>
</tr>
<tr>
<td>TC-REP</td>
<td>-2.00</td>
<td>7.90</td>
<td>-0.67</td>
<td>-3.43</td>
<td>7.50</td>
<td>-1.21</td>
</tr>
<tr>
<td>TC-AI</td>
<td>4.71</td>
<td>8.52</td>
<td>1.46</td>
<td>2.86</td>
<td>9.97</td>
<td>0.76</td>
</tr>
<tr>
<td>Repetitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>-3.71</td>
<td>5.28</td>
<td>-1.86*</td>
<td>-4.00</td>
<td>3.00</td>
<td>-3.54****</td>
</tr>
<tr>
<td>SMT-AI</td>
<td>-3.29</td>
<td>3.20</td>
<td>-2.72***</td>
<td>-4.29</td>
<td>2.98</td>
<td>-3.80******</td>
</tr>
<tr>
<td>TC-REP</td>
<td>0.86</td>
<td>5.01</td>
<td>0.45</td>
<td>0.14</td>
<td>9.15</td>
<td>0.04</td>
</tr>
<tr>
<td>TC-AI</td>
<td>0.43</td>
<td>4.61</td>
<td>0.24</td>
<td>1.00</td>
<td>5.16</td>
<td>0.51</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>-11.00</td>
<td>8.52</td>
<td>-3.42****</td>
<td>-10.29</td>
<td>8.46</td>
<td>-3.22****</td>
</tr>
<tr>
<td>SMT-AI</td>
<td>-6.29</td>
<td>7.87</td>
<td>-2.12**</td>
<td>-4.86</td>
<td>10.96</td>
<td>-1.17</td>
</tr>
<tr>
<td>TC-REP</td>
<td>-1.14</td>
<td>7.88</td>
<td>-0.38</td>
<td>-3.29</td>
<td>12.80</td>
<td>-0.68</td>
</tr>
<tr>
<td>TC-AI</td>
<td>4.29</td>
<td>12.85</td>
<td>0.88</td>
<td>3.86</td>
<td>14.65</td>
<td>0.70</td>
</tr>
</tbody>
</table>

*p < .10
**p < .05
***p < .025
****p < .01
*****p < .005
maintenance (REP, M = -4.86; AI, M = 1.14) was indicated, [F (1, 24) = 3.66, p < .07]. While the training x instruction interaction remained nonsignificant (F < 1), correlated t analyses revealed that the significant within group change for SMT-REP subjects was maintained (see table 3).

Repetition disfluencies. A 2 (SMT vs. TC) x 2 (REP vs. AI) ANOVA for change in self-monitoring errors from SM I to SM II revealed a significant main effect for training [F (1, 24) = 4.57, p < .04] but not for instruction or the training x instruction interaction (F < 1). Evaluation of change score means indicated that subjects receiving SMT manipulations declined in number of self-monitoring errors (M = -3.50) whereas subjects receiving TC manipulations increased in number of Repetition errors committed (M = .22). One-tailed correlated t tests indicated a significant within group change for SMT-AI subjects and a within group trend for subjects receiving SMT-REP manipulations (see table 3).

A 2 x 2 ANOVA for change in Repetition errors from SM I to SM III revealed that SMT (M = -4.15) was still significantly different than TC (M = .57), [F (1, 24) = 4.85, p < .04]. The main effect for instruction and the training x instruction interaction remained nonsignificant (F < 1). Correlated t analyses indicated that the significant within group reduction in self-monitoring errors for SMT-REP and SMT-AI subjects was maintained (see table 3).
Combined disfluencies. A 2 (training) x 2 (instruction) ANOVA for change in self-monitoring errors from SM I to SM II revealed a significant main effect for training \([F (1, 24) = 8.07, p < .01]\) but a nonsignificant effect for instruction \([F (1, 24) = 1.99, p < .17]\) and the training x instruction interaction \((F < 1)\). As expected, subjects receiving SMT manipulations declined in number of "Ah" and Repetition errors committed \((M = -8.65)\), whereas TC subjects increased in number of self-monitoring errors \((M = 1.58)\).

While the main effect for instructions was nonsignificant, change score means indicated that REP subjects \((M = -6.07)\) declined in number of self-monitoring errors more than AI subjects \((M = -1.00)\). Significant within group changes were obtained by subjects receiving both SMT-REP and SMT-AI manipulations (see table 3).

A 2 x 2 ANOVA for change in combined disfluency errors from SM I to SM III indicated that the main effect for training \([F (1, 24) = 3.03, p < .10]\) was no longer significant. The main effect for instruction \([F (1, 24) = 1.94, p < .18]\) and the training x instruction interaction \((F < 1)\) also remained nonsignificant. However, mean reductions in the number of self-monitoring errors on "Ah" and Repetition disfluencies were in the predicted direction (SMT = -7.58, TC = .29, REP = -6.79, AI = -.50). The only within group change which remained significant was for subjects receiving both SMT and REP manipulations (see table 3).
In order to provide a more coherent representation of postmanipulation effects (in terms of reduction in self-monitoring errors on both "Ah" and Repetition disfluencies), change score means are presented in figure 2.

Change in Number of Disfluencies Emitted

The total number of disfluencies (i.e., "Ah's" and Repetitions) emitted by each subject during each self-monitoring period (SM I, II, and III) were acquired directly from the frequency counts made by the criterion rater. In order to test the equivalency of groups prior to the introduction of experimental manipulations, three separate 2 (SMT vs. TC) x 2 (REP vs. AI) ANOVA's were conducted on pre-manipulation (i.e., SM I) disfluency scores. Results revealed no significant main effects or interactions for number of disfluencies (i.e., "Ah's", Repetitions, and both disfluencies combined) emitted (smallest $p > .20$). Pre-manipulation means for number of emitted disfluencies (both "Ah's" and Repetitions) were as follows: SMT-REP ($M = 30.43$), SMT-AI ($M = 29.00$), TC-REP ($M = 41.43$), and TC-AI ($M = 32.29$).

In order to evaluate change in number of disfluencies from pre-manipulation (SM I) to post-manipulation periods (SM II and SM III), disfluency scores (for SM I, II, and III) were converted into two separate change scores. As in previous error analyses, the first score represented change in number of disfluencies emitted from SM I to SM II, and
Figure 2. Change in mean number of self-monitoring errors (for "Ah's" and Repetitions combined) from pre-manipulation period (SM I).
the second score represented change in disfluencies emitted from SM I to SM III.

In order to simplify the analysis of change scores, an initial assessment of therapist differences was conducted. The resulting 3 (therapists) x 2 (change score) repeated measures ANOVA for unequal n revealed no significant differences between therapists on change in number of disfluencies emitted by their respective group members. These results were consistent for "Ah's", Repetitions, and both disfluency categories combined (smallest p > .20).

Because of the nonsignificant effect for therapists, this factor was eliminated in all subsequent analyses. Change score means for each experimental condition are presented in table 4.

"Ah" disfluencies. A 2 (training) x 2 (instruction) ANOVA for change in "Ah" disfluencies emitted from SM I to SM II revealed a nonsignificant main effect for training and a nonsignificant training x instruction interaction (both Fs < 1). Results also indicated a nonsignificant trend for instructions [F (1, 24) = 3.49, p < .08], suggesting that REP subjects (M = -9.21) had made a sharper reduction in "Ah" disfluencies than had AI subjects (M = -3.14). One-tailed correlated t analyses revealed significant within group reductions in "Ah" disfluencies for subjects receiving both SMT-REP and TC-REP manipulations (see table 4).
## TABLE 4

CHANGE IN MEAN NUMBER OF EXTERNALLY-MONITORED DISFLUENCIES FROM PRE-MANIPULATION PERIOD

**(SM I)**

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>SM II</th>
<th>SM III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>&quot;Ah's&quot;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>-10.00</td>
<td>10.41</td>
</tr>
<tr>
<td>SMT-AI</td>
<td>-2.29</td>
<td>6.26</td>
</tr>
<tr>
<td>TC-REP</td>
<td>-8.43</td>
<td>7.30</td>
</tr>
<tr>
<td>TC-AI</td>
<td>-4.00</td>
<td>9.73</td>
</tr>
<tr>
<td>Repetitions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>-2.71</td>
<td>9.20</td>
</tr>
<tr>
<td>SMT-AI</td>
<td>-3.86</td>
<td>5.64</td>
</tr>
<tr>
<td>TC-REP</td>
<td>-3.14</td>
<td>4.56</td>
</tr>
<tr>
<td>TC-AI</td>
<td>-3.71</td>
<td>6.87</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>-12.71</td>
<td>17.28</td>
</tr>
<tr>
<td>SMT-AI</td>
<td>-6.14</td>
<td>10.46</td>
</tr>
<tr>
<td>TC-REP</td>
<td>-11.57</td>
<td>5.13</td>
</tr>
<tr>
<td>TC-AI</td>
<td>-7.71</td>
<td>15.73</td>
</tr>
</tbody>
</table>

*P < .05
**P < .025
***P < .01
****P < .005
*****P < .0005
A 2 x 2 ANOVA for change in "Ah" disfluencies emitted from SM I to SM III revealed nonsignificant main effects for both training and instruction and a nonsignificant training x instruction interaction (smallest $p > .20$). Correlated $t$ analyses indicated that significant within group changes for SMT-REP and TC-REP subjects had been maintained (see table 4).

In order to evaluate nonspecific reduction in "Ah" disfluencies over time, group data were combined and a one-way repeated measures ANOVA on the actual number of "Ah's" emitted across SM I, SM II, and SM III was conducted. Results indicated a significant decline in number of "Ah" disfluencies emitted across self-monitoring periods ($F(2, 54) = 10.79, p < .0003$). Newman-Keuls post hoc comparisons showed that the mean number of "Ah" disfluencies emitted at SM I ($M = 20.86$) was significantly different than the number emitted at both SM II ($M = 14.68$) and SM III ($M = 12.96$), ($p < .01$). However, the mean number of "Ah's" at SM II did not differ significantly from the number emitted at SM III.

Repetition disfluencies: A 2 (training) x 2 (instruction) ANOVA for change in Repetition disfluencies emitted from SM I to SM II revealed nonsignificant main effects for training and instruction and a nonsignificant training x instruction interaction (all $F$s $< 1$). One-tailed correlated $t$ tests indicated no significant within group reductions in
number of Repetitions emitted (see table 4).

A $2 \times 2$ ANOVA for change in Repetition disfluencies from SM I to SM III also revealed nonsignificant main effects and interactions (all $F_s < 1$). However, analyses of within group change indicated that subjects receiving SMT-AI manipulations declined significantly in number of Repetitions emitted from SM I to SM III (see table 4).

The one-way repeated measures ANOVA conducted on actual number of Repetition disfluencies emitted across SM I, SM II, and SM III indicated a significant decline in disfluencies across self-monitoring periods [$F(2, 54) = 5.56, p < .007$]. Subsequent Newman-Keuls post hoc comparisons revealed that the mean number of Repetitions emitted at SM I ($M = 12.43$) was significantly different than the number emitted at both SM II ($M = 9.07$) and SM III ($M = 8.89$), ($p < .05$), but that no difference existed between the number of Repetitions emitted at SM II and SM III.

**Combined disfluencies.** A $2 \times 2$ (training) x 2 (instruction) ANOVA for change in combined disfluencies emitted (i.e., both "Ah's" and Repetitions) from SM I to SM II revealed nonsignificant main effects for training and instruction and a nonsignificant training x instruction interaction (smallest $p > .30$). However, within group correlated $t$ analyses indicated that subjects receiving SMT-REP and TC-REP manipulations declined significantly in number of combined disfluencies emitted (see table 4).
A second 2 x 2 ANOVA for change between SM I and SM III again revealed nonsignificant main effects and interactions (all $F_{s} < 1$). One-tailed correlated $t$ analyses indicated that the within group change in disfluencies emitted from SM I to SM III remained significant only for subjects receiving TC-REP manipulations (see table 4).

The one-way ANOVA on actual number of combined disfluencies emitted across SM I, SM II, and SM III revealed a significant decline across self-monitoring periods [$F (2, 54) = 11.64, p < .0002$]. Newman-Keuls post hoc comparisons showed that the mean number of "Ah" and Repetition disfluencies emitted at SM I ($M = 33.29$) was significantly greater than the number emitted at both SM II ($M = 23.75$) and SM III ($M = 21.86$), ($p < .01$), although no differences existed between disfluencies emitted at SM II and SM III.

Change score means for "Ah's" and Repetitions combined are shown in figure 3.

**Percentage of "Accurately" Monitored Speeches**

In order to corroborate the results of previous error score analyses, data were compiled on the percentage of speeches in which either perfect agreement or a one disfluency deviation existed between the self-recorder and the criterion rater. Group percentages were based on 21 speeches per self-monitoring period (i.e., 3 speeches per subject) and comparisons between the number of disfluencies reported
Figure 3. Change in mean number of externally-monitored disfluencies ("Ah's" and Repetitions combined) from pre-manipulation period (SM I).
by self-recorders and the number reported by the rater were made on "Ah's", Repetitions, and both disfluencies combined. Percentages for each experimental condition were obtained by counting the number of speeches meeting the zero to one deviation criterion and then dividing by the number of speeches given during each self-monitoring period (i.e., twenty-one). Percentage data are presented in table 5.

The percentages shown in this table tend to corroborate previous error score data in that SMT-REP and SMT-AI manipulations precipitated increases in the percentage of "accurately" monitored speeches at post-manipulation periods (i.e., SM II and III). Although a decline in the percentage of speeches meeting the zero to one accuracy criterion on "Ah" disfluencies occurred between SM II and SM III for SMT-REP and SMT-AI subjects, percentage data for Repetitions and combined disfluencies indicated that accuracy was either maintained or increased at SM III. However, one-tailed $z$ tests for the significance of a proportion (see Bruning & Klintz, 1968) indicated that significant within group changes were exhibited only by subjects receiving SMT-REP manipulations (see table 5). It should be noted that the only nonsignificant within group change for the SMT-REP condition occurred at SM III on "Ah" disfluencies.

Data on the percentage of speeches in which perfect agreement existed between self-recorders and the criterion rater correspond closely with the data presented in table 5.
### TABLE 5

PERCENTAGE OF OBSERVATIONS (SPEECHES) WITH PERFECT OR ONE DISFLUENCY DEVIATION BETWEEN SELF-RECORDERs AND CRITERION RATER

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Self-monitoring period</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SM I</td>
<td>SM II</td>
<td>SM III</td>
<td></td>
</tr>
<tr>
<td>&quot;Ah's&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>33</td>
<td>62*</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>SMT-AI</td>
<td>33</td>
<td>48</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>TC-REP</td>
<td>38</td>
<td>33</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>TC-AI</td>
<td>52</td>
<td>29</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Repetitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>29</td>
<td>62*</td>
<td>67*</td>
<td></td>
</tr>
<tr>
<td>SMT-AI</td>
<td>38</td>
<td>48</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>TC-REP</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>TC-AI</td>
<td>43</td>
<td>48</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>31</td>
<td>62*</td>
<td>62*</td>
<td></td>
</tr>
<tr>
<td>SMT-AI</td>
<td>36</td>
<td>48</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>TC-REP</td>
<td>31</td>
<td>29</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>TC-AI</td>
<td>48</td>
<td>38</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

Note. Percentages based on 21 speeches per self-monitoring period (i.e., 3 speeches per subject).

*p < .05, one-tailed z test for significance of correlated proportions (Bruning & Klíntz, 1968) using SM I as expected value.
As an indication, the percentages of speeches meeting the zero deviation criterion at SM III for combined disfluencies were as follows: SMT-REP (38%), SMT-AI (31%), TC-REP (14%), TC-AI (14%).

Correlations Between Self-recorders and Criterion Rater

Another method of analyzing the accuracy of the self-monitored data collected during the course of the present investigation was accomplished by correlating subject scores with criterion rater scores. Table 6 shows the mean correlations between the number of disfluencies ("Ah's" and Repetitions combined) recorded by subjects and the number recorded by the criterion rater. In order to maintain independence of measures, separate correlations were computed for each of the three speeches given during each self-monitoring period and then averaged via z scores (see Edwards, 1950). Mean correlations (r) and coefficients of determination (r²) are presented for both interactions and main effects across pre-manipulation (SM I) and post-manipulation periods (SM II, SM III).

The correlations presented in table 6 (i.e., for group interactions) indicate that the agreement between self-recorders and the criterion rater increased at post-manipulation (SM II) for subjects receiving SMT-REP and SMT-AI procedures. Moreover, the coefficient of determination for the SMT main effect reveals that the correlation
### TABLE 6

MEAN CORRELATIONS BETWEEN THE NUMBER OF SELF- AND EXTERNALLY-MONITORED SPEECH DISFLUENCIES ("AH'S" AND REPETITIONS)

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Self-monitoring period</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SM I</td>
<td>SM II</td>
<td>SM III</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>r²</td>
<td>r</td>
<td>r²</td>
<td>r</td>
</tr>
<tr>
<td>Interactions^a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT-REP</td>
<td>.56</td>
<td>.31</td>
<td>.71</td>
<td>.50</td>
<td>.68</td>
</tr>
<tr>
<td>SMT-AI</td>
<td>.44</td>
<td>.19</td>
<td>.71</td>
<td>.50</td>
<td>.66</td>
</tr>
<tr>
<td>TC-REP</td>
<td>.20</td>
<td>.04</td>
<td>.17</td>
<td>.03</td>
<td>.29</td>
</tr>
<tr>
<td>TC-AI</td>
<td>.60</td>
<td>.36</td>
<td>.18</td>
<td>.03</td>
<td>.40</td>
</tr>
<tr>
<td>Main Effects^b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>.50</td>
<td>.25</td>
<td>.71</td>
<td>.50</td>
<td>.67</td>
</tr>
<tr>
<td>TC</td>
<td>.39</td>
<td>.15</td>
<td>.18</td>
<td>.03</td>
<td>.34</td>
</tr>
<tr>
<td>REP</td>
<td>.39</td>
<td>.15</td>
<td>.49</td>
<td>.24</td>
<td>.51</td>
</tr>
<tr>
<td>AI</td>
<td>.53</td>
<td>.28</td>
<td>.49</td>
<td>.24</td>
<td>.54</td>
</tr>
</tbody>
</table>

^a n = 7

^b n = 14 (Collapsed across interacting conditions)
at SM II \( (r = .71) \) is exactly twice as large as the correlation at SM I \( (r = .50) \). Despite the fact that the sample size was quite small, a one-tailed \( z \) test on the SMT \( (r = .71) \) and TC \( (r = .18) \) correlations at post-intervention (SM II), revealed that the two differed significantly \( (p < .05, \text{ one-tailed}) \). It is also apparent from table 6 that the correlations at SM III decline for subjects receiving SMT manipulations. However, despite the lack of maintenance in self-monitoring accuracy, correlations at SM III for SMT subjects remain appreciably larger than those for TC subjects.

**Anxiety Reduction Measures**

In order to evaluate change in self-report of public-speaking anxiety, analyses were conducted on pre-post PRCS scores and on mean Likert-scale anxiety ratings (taken at SM I, II, and III). Assessment of anxiety reduction data also allowed for the evaluation of differential effects as a result of training and instruction manipulations.

**PRCS questionnaire.** A 2 (SMT vs. TC) x 2 (REP vs. AI) repeated measures (pre-post) ANOVA on PRCS scores revealed no significant differences between groups, interactions, or group x pre-post interactions (smallest \( p > .20 \)). However, the repeated measures (pre-post) main effect was highly significant \( [F (1, 24) = 148.19, p < .0001] \), indicating a substantial decline in self-report of public speaking anxiety from pre- \( (M = 24.71) \) to post-treatment \( (M = 13.00) \).
Mean Likert-scale ratings. A 2 (SMT vs. TC) x 2 (REP vs. AI) repeated measures (SM I, II, and III) ANOVA conducted on mean (11-point) Likert-scale ratings again revealed no significant differences between groups, interactions, or group x repeated measures interactions (smallest $p > .20$). However, the repeated measures main effect did attain significance \( F(2, 48) = 70.51, p < .0001 \). Subsequent Newman-Keuls post hoc comparisons indicated that consistent and statistically significant ($p < .01$) declines in perceived anxiety took place over the three public-speaking sessions (SM I, $M = 6.05$; SM II, $M = 4.71$; SM III, $M = 2.77$).

Post-experimental Questionnaire

In order to gain further information from subjects regarding certain aspects of the present investigation, the Program Evaluation Questionnaire (see appendix H) was administered immediately after the final set of speeches had been completed (i.e., following SM III). The results of this seven-item questionnaire were as follows:

(a) While 29 percent of all subjects reported that they would not have participated in the speech anxiety treatment program if they had not received experimental credit, 32 percent reported that they would have participated without this incentive, and 39 percent were not sure. These data offer moderate support for the external validity of results obtained in the current investigation in that a large percentage of
subjects volunteered for reasons relevant rather than irrelevant to treatment.

(b) When subjects were asked to explain what was being studied during the current treatment program, 86 percent mentioned speech anxiety, 7 percent mentioned speech disfluencies and anxiety, and 7 percent stated that public speaking ability was being investigated. None of the subjects cited reliability or accuracy of self-monitoring as being of empirical interest.

(c) Although three REP subjects questioned the utility of the personality profile feedback, none of the remaining participants reported being suspicious of any other portion of the speech anxiety program. It would therefore appear that subjects were not aware that their speeches were being externally-monitored and that the observation process was in fact unobtrusive.

(d) Subjects (on the average) reported that they expected to experience a moderate decline in subjective anxiety during their final three speeches (i.e., at SM III). The overall mean expectation was rated 2.39 on a 0 to 4 point scale (i.e., 0 = no decline whatsoever, 2 = moderate decline, 4 = dramatic decline). A 2 (training) x 2 (instruction) ANOVA conducted on responses to this item revealed nonsignificant differences (smallest p > .25). Since this item served as a check on the effect produced by the demand characteristics manipulation employed prior to SM III (see p. 49), it
would appear that expectation for behavior change was only "moderately" influenced.

(e) Subjects (on the average) also reported that they expected a moderate decline in objective anxiety (i.e., number of "Ah's" and Repetitions) at SM III. The overall mean for this item was 2.14 on the same 0 to 4 point scale. While the 2 x 2 ANOVA conducted on responses to this item yielded nonsignificant differences (smallest p > .20), the overall mean expectation again suggests only "moderate" impact for the demand characteristic manipulation.

(f) When asked how accurately "Ah's" and Repetitions were monitored during speech presentations, subjects (on the average) reported being moderately accurate (i.e., M = 6.29 on a 0 to 10 point Likert-type scale). A 2 x 2 ANOVA conducted on responses to this item revealed that REP subjects (M = 7.18) felt they were more accurate in monitoring disfluencies than were AI subjects (M = 5.39), [F (1, 24) = 6.21, p < .02].

(g) In relation to the previous item, subjects were also asked to rate how hard they tried to accurately self-monitor. On a similar 0 to 10 point Likert-type scale, the overall mean was 7.82, indicating that subjects attempted to monitor accurately even though they felt they were only "moderately" successful. A 2 x 2 ANOVA conducted on responses to this item again indicated that REP subjects (M = 8.36) felt they tried harder to monitor reliably than
did AI subjects \( (M = 7.29) \), \( F (1, 24) = 3.99, p < .06 \).

Results of the statistical analyses on the two post-experimental items regarding accuracy offer support for the validity of the REP procedures employed.
CHAPTER IV

DISCUSSION

The intent of the present investigation was to create a clinically-relevant analogue situation where two diverse procedures, aimed at increasing the accuracy of self-monitored data, could be empirically evaluated. The first procedure, Self-monitoring Training (SMT), was designed to increase the initial level of self-monitoring accuracy by (a) allowing subjects to gain practice in external-monitoring, (b) providing feedback as to the accuracy of externally-monitored responses, (c) socially-reinforcing accurate data, and (d) transferring subjects to a self-monitoring situation where accuracy-feedback was made available. The entire SMT procedure was strategically graduated in that subjects progressed from external-monitoring to self-monitoring, from one target behavior to two target behaviors, and from monitoring periods of 15 seconds to monitoring periods of up to 3 minutes. The second procedure, Reliability Enhancement Package (REP), was designed to decrease demand characteristics for behavior change by employing (a) cognitive consistency, (b) consequence clarification, and (c) cueing manipulations. The primary purpose of
REP was to create the overt and persistent suggestion that veridical self-monitoring rather than demonstration of treatment efficacy was the important task at hand.

To more thoroughly evaluate these two procedures, SMT was compared to a Training Control (TC) condition while REP was evaluated against an Accuracy Instruction (AI) condition. The TC procedure was designed to control for exposure to irrelevant training stimuli in order for experimental effects to be attributable to the accuracy enhancing characteristics specific to SMT. The AI control was designed to test REP against a more elementary procedure composed of accuracy instructions (adapted from Jeffrey, 1974b; Taplin & Reid, 1973) and a less elaborate cueing statement (see Edelstein & Noah, Note 11).

The above-mentioned procedures were evaluated within a clinical context where speech anxious undergraduates, who were participating in a speech anxiety treatment program, self-monitored speech disfluencies in order to aid therapists in evaluating the efficacy of the in vivo flooding procedure utilized. Self-monitored data were collected during three separate in vivo flooding sessions both prior to (i.e., at SM I) and following (i.e., at SM II and SM III) the administration of experimental procedures.

The results of the present investigation indicated that (a) SMT was more effective than TC in increasing the accuracy of self-monitored data, (b) REP as a sole accuracy-
enhancing manipulation was more effective than AI only when the target behavior was easily discriminable, and (c) a combination of SMT and REP procedures proved most effective in improving the accuracy of self-monitored data and in maintaining these gains over time. A more comprehensive discussion of present results follow.

Self-monitoring Training

Four days following the administration of experimental procedures (i.e., at SM II), SMT was found to be significantly more effective than TC in reducing the number of self-monitoring errors on all target behaviors (i.e., "Ah" disfluencies, Repetitions, and both categories combined). However during the third self-monitoring session (which occurred eight days following the administration of experimental procedures and four days following the manipulation of demand characteristics), SMT differed from TC only in reductions on Repetition errors, although a between groups trend was indicated for both disfluency categories combined. The fact that differences between SMT and TC were not maintained at SM III was primarily the result of the decline in "Ah" disfluency errors made by TC-REP subjects and the increase in "Ah" disfluency errors made by SMT-AI subjects between SM II and SM III. As will be discussed more thoroughly later, REP manipulations appear to be highly important in affecting reductions in "Ah" disfluency errors and in maintaining these gains over time.
Corresponding with results on reductions in self-monitoring errors, both SMT conditions (i.e., SMT-REP and SMT-AI) increased in percentage of speeches with zero or one disfluency deviation between self-monitored and externally-monitored (i.e., criterion-rated) data. These increased percentages of "accurately" monitored speeches were maintained eight days following the administration of training and showed 59 percent of SMT speeches as compared to 36 percent of TC speeches to be monitored "accurately." Additionally, the correlation between data gathered by the criterion rater doubled between pre- and post-manipulation periods (i.e., $r = .50$ at SM I, $r = .71$ at SM II), and declined only slightly four days later (i.e., $r = .67$ at SM III). These inter-rater correlations were appreciably greater than those achieved by TC subjects (i.e., $r = .18$ at SM II, $r = .34$ at SM III).

**Reliability Enhancement Package**

While REP in conjunction with Self-monitoring Training produced consistent and significant within group declines in number of monitoring errors committed, REP as a sole accuracy modifying manipulation was not uniformly effective. Although results indicated that REP produced significantly greater reductions in "Ah" disfluency errors than did the AI condition, differences between groups on reductions in Repetition errors did not achieve statistical significance. However, recent speech disfluency literature suggests that
subjects do not readily perceive the occurrence of Repetitions but easily recognize the emission of interjection (i.e., "Ah") disfluencies (e.g., Brutten & Shoemaker, 1967, 1971; Oelschlaeger & Brutten, 1976). Data from the present investigation tend to support this finding in that the overall mean correlation between subjects and the criterion rater at pre-manipulation was .61 for "Ah" disfluencies and .24 for Repetitions. It may therefore be that REP is only effective in those cases where the target behavior is easily discriminable (e.g., "Ah" disfluencies), whereas in situations where the occurrence of the target behavior is less easily perceived (e.g., Repetition disfluencies), REP may add little to the accuracy-enhancing effects produced by Self-monitoring Training. The above is supported by within group error reduction data which indicates that both SMT groups (i.e., SMT-REP, SMT-AI) showed significant within group declines in Repetition errors while only SMT-REP made significant reductions in self-monitoring errors on "Ah" disfluencies.

It was also originally hypothesized that neither REP or AI would be effective in reducing self-monitoring errors at SM II, although following the manipulation of demand characteristics at SM III, REP would serve to maintain the gains in accuracy demonstrated at SM II while AI subjects would increase in the number of self-monitoring errors committed (i.e., as a result of treatment demand). This
hypothesis received partial support in that subjects receiving SMT-REP manipulations maintained significant reductions in self-monitoring errors at SM III whereas SMT-AI subjects did not. However, results also indicated that REP did not affect monitoring errors solely at SM III. In fact, REP and AI differed significantly on reductions in "Ah" disfluency errors at SM II while only a trend toward maintenance was apparent at SM III. Moreover, SMT-REP subjects showed significant within group declines on "Ah" disfluency errors at SM II whereas SMT-AI subjects did not. Both of these results run counter to the original hypothesis that REP would serve only to maintain gains in self-monitoring accuracy at SM III by reducing the biasing effects of demand characteristics. Two possible explanations for the lack of support for this hypothesis are as follows: (a) The impact of the demand manipulation employed between SM II and SM III was relatively weak in that subjects reported only moderate expectations for behavior change (i.e., reductions in disfluencies) at SM III. Although subject expectations were not assessed prior to SM II, it may well be that demand for behavior change was not appreciably altered as a result of the instructional manipulation employed and that treatment demand was as strong at SM II as at SM III. In fact, it could be argued that demand may have been lower at SM III in view of the actual reduction in disfluencies which took place at SM II. In either case, REP may have affected self-monitoring accuracy
at SM II by reducing the impact of the demand characteristics which existed at that time. (b) REP may not affect accuracy of self-monitoring solely by minimizing the impact of demand characteristics and may instead simply sensitize subjects to the importance of accurate data so that more energy is invested in self-monitoring responsibilities. This explanation is supported by post-experimental reports that REP subjects felt they tried harder to accurately self-monitor than did AI subjects. While more research is needed to explore the manner in which REP affects self-monitoring accuracy, it is indeed plausible that accuracy may have been enhanced by making subjects more aware of the importance of self-reported data (see McFall, Note 13).

Procedural Combinations

Clearly the most effective combination of accuracy-enhancing strategies employed during the present investigation was composed of both SMT and REP procedures. Subjects receiving the SMT-REP procedural combination demonstrated significant within group declines in self-monitoring errors on all target behaviors and maintained these gains in accuracy eight days following the administration of experimental procedures. Moreover, results indicated that SMT-REP subjects were able to eliminate over 50 percent of their self-monitoring errors following the presentation of training and instruction manipulations.

Corroborating data also indicated that the percentage
of speeches monitored "accurately" (i.e., 0 to 1 disfluency deviation with criterion rater) doubled between pre- (SM I) and post-manipulation (SM II) speech sessions and that this percentage increase was maintained over time (i.e., at SM III). Additionally, the correlation between the self-recordings of SMT-REP subjects and criterion ratings increased from .56 at pre-manipulation to .71 at post-manipulation (SM II). This increased relationship between self- and externally-monitored data also showed maintenance characteristics at SM III (i.e., $r = .68$).

It should be emphasized that the level of correspondence ($r = .71$) between self-monitored and externally-monitored (i.e., criterion-rated) data for subjects in both SMT conditions is quite impressive when consideration is given to the situational context in which self-monitoring took place. First of all, subjects monitored verbal responses in an environmental setting which was highly anxiety provoking. More specifically, pre-treatment PRCS scores indicate that the level of public speaking anxiety for subjects in the present investigation was comparable to or even higher than the level exhibited by participants in previous speech anxiety studies (e.g., Kirsch, Wolpin, & Knutson, 1975; Meichenbaum, Gilmore, & Fedoravicious, 1971; Paul, 1966). Secondly, subjects were allowed minimal rehearsal time prior to speech presentations and were therefore required to expend a substantial amount of energy organizing topic-relevant thoughts.
during actual speech delivery. Thirdly, subjects were re-
quired to monitor two verbal behaviors during the presenta-
tion of speeches. Prior research has demonstrated that
self-monitoring accuracy is sharply reduced when a second
concurrent operant task is introduced (Epstein et al.,
1975, 1976). In light of these factors, obtained inter-
rater correlations above .70 would appear to magnify the
methodological significance of the present results. More-
over, the accuracy-enhancing qualities of SMT-REP manipula-
tions would, no doubt, be even more powerful in situations
where competing stimuli are not so pervasive. However, more
research is clearly necessary in order to substantiate this
claim.

Before progressing further, it should be noted that
TC-AI subjects increased in mean number of self-monitoring
errors (see figure 2) while all other groups declined in
the number of errors committed. Although these subjects re-
ceived accuracy instructions, a cue reminding them to "re-
cord," and were exposed to training stimuli, they were not
administered procedural components specifically designed to
increase self-monitoring accuracy. This finding would
appear to be related to results of previous studies con-
cerning external-monitoring where declines in accuracy
were noted as a result of instrument decay (e.g., Reid,
1970; Romanczyk et al., 1973). In the present investigation,
subjects were given response definition sheets, prior to
their initial speeches, which were studied until each therapist was confident that target behaviors could be reliably identified. The increase in monitoring errors made by AI-TC subjects from pre- (SM I) to post-manipulation periods is therefore most likely due to the fact that (a) subjects did not receive SMT procedures which would have served to recalibrate self-observers and to train-up self-monitoring skills, and (b) subjects did not receive manipulations designed to repeatedly emphasize the importance of honest and accurate self-recorded data. This finding should also serve to warn applied researchers against assuming that self-monitoring accuracy will improve as a result of decrements in maladaptive target behaviors. Present results indicate that for subjects receiving control conditions, self-monitoring errors increased despite reductions in the frequency of target responses (see figure 3).

Therapeutic Effects

The relevance of the present clinical analogue to the regular treatment setting is enhanced by the fact that the speech anxiety treatment program, which served as a medium to study self-monitoring accuracy, was actually quite effective. Results revealed (a) significant reductions in public speaking anxiety (as indicated by pre-post PRCS scores and mean Likert anxiety ratings) and (b) significant declines in the number of speech disfluencies emitted (i.e., for both
"Ah" and Repetition disfluencies). Moreover, the decline in self-reported anxiety (as revealed by pre- to post-treatment PRCS scores) indicates that treatment effects in the present investigation are comparable to the effects shown in previous studies where the modification of public speaking anxiety was the main focus of empirical interest and where more elaborate treatment strategies (i.e., cognitive self-instruction, systematic desensitization) were employed (e.g., Meichenbaum et al., 1971).

While it is quite tempting to attribute reduction in subjective anxiety to the therapeutic effects of repeated exposure (Kirsch et al., 1975) and decline in disfluency production to self-monitoring (Cavior & Marabotto, 1976), multiple treatment interference (Campbell & Stanley, 1966) prevents this interpretation. More specifically, either one or both of these treatment effects could have been the result of a number of interacting factors (e.g., repeated exposure, treatment rationale administration, self-monitoring, demand characteristics, etc.). Since the intent of the present investigation was to evaluate procedures aimed at increasing self-monitoring accuracy rather than to identify the cause(s) of therapeutic change, these results must yet await further empirical clarification.

However, in that anxiety reduction and declines in emitted disfluencies may have affected self-monitoring accuracy, therapy effects become quite relevant to the
purposes of the present study. In this regard, it could be potentially argued that self-monitoring accuracy improved as a result of anxiety reduction and target behavior decrement rather than as a result of training and instruction. However, results of the present investigation indicated that experimental conditions did not differ on indices of anxiety reduction or on declines in disfluencies emitted while differing significantly on reductions in self-monitoring errors. If increased accuracy covaried strictly with decrements in these two variables, between group differences would be expected. Nevertheless, results did reveal significant within group reductions in disfluencies emitted for particular experimental conditions. As can be seen in figure 3, both SMT-REP and TC-REP subjects demonstrated significant declines in mean number of disfluencies emitted at SM II. However, figure 2 indicates that only SMT-REP subjects were able to take advantage of the disfluency decline and show comparable reductions in self-monitoring errors. Similarly, both SMT-AI and TC-AI subjects showed corresponding decrements in emitted disfluencies (see figure 3), while only SMT-AI subjects were able to demonstrate significant reductions in monitoring errors (see figure 2). Although a relationship between anxiety reduction, target behavior decrement, and the accuracy of self-monitoring cannot be totally discounted, the correlation between these variables does not appear to be strong enough to threaten the internal validity of the
present results.

Methodological Implications

During the past few years, self-monitoring has become an increasingly popular assessment tool and has been employed in a variety of clinical settings to evaluate change in a wide range of problem behaviors (Kazdin, 1974; Thoresen & Mahoney, 1974). While the rise in popularity of self-monitoring has paralleled the growth in self-control or self-management approaches to treatment, other explanations exist for the increased interest in this particular applied assessment strategy. In this regard, McFall (Note 13) has recently suggested several reasons why self-monitoring as opposed to external-monitoring may be the preferred method of evaluation in many clinical situations where behavioral assessment is deemed appropriate: (a) self-monitoring methods are cost-efficient in that clients, rather than paid trained observers are employed in the data gathering process, (b) self-monitoring may be the only practical and ethical procedure available for obtaining detailed information on particular forms of "sensitive, private, or inaccessible behavior", (c) self-monitoring procedures minimize the undesirable effects created by the presence of obtrusive observers, and (d) self-monitoring makes it possible for applied practitioners to obtain dense coverage of problem behavior occurrences in that self-recordings can be obtained
in a wide variety of environmental settings. While the advantages stemming from the use of self-monitoring are quite convincing, research-oriented practitioners have been reluctant to abandon external-observation in favor of an assessment strategy which lacks the methodological rigor and objectivity of external-monitoring procedures. Indeed, numerous studies attest to the unreliability of behavioral observations gathered by clients who were enlisted as active collaborators in the data gathering process (e.g., Broden et al., 1971; Fixsen et al., 1972; McFall, 1970).

In view of reasons favoring the use of self-monitoring for assessment purposes and because of the problems posed by the unreliability of such behavioral records, several attempts have been made to develop strategies to increase the accuracy of self-recorded data (e.g., Bolstad & Johnson, 1972; Drabman et al., 1973; Epstein et al., 1976; Fixsen et al., 1972; Lipinski et al., 1975; Risley & Hart, 1968; Turkowitz et al., 1975). While many of these procedures have proved successful in improving the correspondence between self-monitored and externally-monitored data, each of these intervention strategies have relied upon tangible rewards to motivate subjects to become more reliable in self-monitoring responsibilities. The major problem with reinforcement-related interventions is that the practitioner must know when the client is and is not producing accurate data so that the contingent relationship between rewards and
reliability can be maintained. This would appear to defeat the original intention of self-monitoring in that the client is no longer independently responsible for data collection (Bandura, 1971; Buckley, 1968; Glynn & Thomas, 1974; Kanfer, 1971). Moreover, if external-observation is necessary to corroborate data collected by the self-observer, it would appear that self-monitoring is superfluous since the practitioner would most likely want to employ data from the most reliable source (i.e., from trained observers) to evaluate treatment outcome. Two additional problems with reinforcement oriented procedures include: (a) it is not possible to reinforce accuracy when target behaviors are covert or when overt behaviors must be assessed in multiple settings, and (b) self-monitoring accuracy rapidly deteriorates when externally-imposed incentives are withdrawn.

The results of the present investigation indicate that alternative means of improving self-monitoring accuracy are available to applied researchers which make it unnecessary to arrange for continuous external reliability assessment. More specifically, Self-monitoring Training and Reliability Enhancement procedures are pre-monitoring manipulations which make it possible to increase the accuracy of self-monitoring before the data collection process is actually initiated. Moreover, SMT and REP procedures can also be periodically employed after self-monitoring has commenced in order to recalibrate self-observers and to reemphasize the importance of accurate behavioral records.
The current procedures are also quite flexible in terms of the level of confidence desired by empirically-oriented practitioners in the data obtained from self-observers. More specifically, in innovative case studies where methodological rigor can be partially relaxed (Jeffrey, 1974a; Lazarus & Davison, 1971), SMT and REP procedures may prove to be expedient methods of enhancing self-report accuracy. While it would be unrealistic to have total confidence in the reliability and validity of self-reported data even after these accuracy-enhancing manipulations had been employed, the probability of obtaining reliable self-reports would appear greatly improved. In situations where greater objectivity is desired, covert reliability checks could be performed in order to assess self-monitoring accuracy after SMT and REP procedures had been instigated. Data obtained from such spot checks would serve to heighten the confidence in the self-monitored data obtained and would also serve to inform the applied researcher when recalibration or reemphasis on self-report accuracy was needed.

Self-monitoring Training and Reliability Enhancement procedures would also appear to have applicability when covert behaviors (i.e., thoughts, hallucinations, etc.) become targets for treatment interventions. In this regard, thoughts and covert self-statements could be modeled overtly and the client trained to reliably discriminate target statements from non-target statements via systematic SMT methods.
In fact, a recent editorial note reported that modeling and rehearsal-training procedures have already been used to aid chronic schizophrenics, adult psychiatric outpatients, male alcoholics, and college students in the observation and reporting of covert self-instructional behaviors (Meyers, Mercatoris, & Artz, 1976). Furthermore, REP manipulations could be additionally employed to emphasize the importance of accurate monitoring and to "cue-off" self-recording behaviors.

One primary problem which prevents the application of REP (i.e., as it was employed in the current research) in situations involving clients seeking mental health services is the deception involved in the cognitive consistency component of REP. This particular manipulation was included as part of the Reliability Enhancement Package because of previous research demonstrating that subjects tend to act in accordance with interpretive feedback (i.e., "bogus" personality attributes) from personality inventory profiles (e.g., Aronson & Mettee, 1968; Graf, 1971). It was hypothesized that a similar effect might occur with self-monitored data if subjects were led to believe that their trait of honesty and accuracy were of particular importance in obtaining an accurate evaluation of the speech anxiety treatment program. While no attempt was made to parcel out the reliability enhancing effects produced solely by this one experimental manipulation, REP
would have to be modified to eliminate the "bogus" nature of the cognitive consistency component before it could be applied in clinical settings. However, the other two REP components (i.e., consequence clarification and cueing statement) are not prone to the deception criticism and can be employed without ethical constraints in clinically-oriented situations.

In closing, self-monitoring would appear to hold much promise as a behavioral assessment strategy for use in applied clinical settings. While self-monitoring methods have already been employed in the evaluation of over 28 clinically-related behaviors (McFall, Note 13), the problem of self-monitoring accuracy has remained a major stumbling block to the wide acceptance of self-assessment procedures. Attempts to overcome the obstacle of self-monitoring inaccuracy have made use of primary rewards (Epstein et al., 1976; Lipinski et al., 1975; Risley & Hart, 1968), matching and fading procedures (Bolstad & Johnson, 1972; Drabman et al., 1973; Turkewitz et al., 1975), stimulus cues (Edelstein & Noah, Note 11) and procedural packages (Bornstein et al., Note 12) in order to increase the reliability of self-observational procedures. In view of the favorable results shown in the present research, self-monitoring training may now be added to this rapidly accumulating list. While problems specific to each of the procedures mentioned above are quite apparent, the diversity of these accuracy-enhancing
strategies may, in the near future, lead to combinations of techniques which are maximally effective in particular assessment situations. Although the scientific method encourages component analyses to tear down procedural packages in order to point out specific cause-effect relationship, it is recommended that current research be directed toward the development of multi-faceted procedures which not only increase self-monitoring accuracy but also maintain high levels of reliability over prolonged periods of time. In view of the potential benefits to be derived from accurate and reliable self-monitoring methods, creative research in this area would appear to have major implications for behavioral psychology.
CHAPTER V

SUMMARY

The intent of the present investigation was to create a clinically-relevant analogue situation where two diverse procedures, aimed at increasing the accuracy of self-monitored data, could be empirically evaluated. More specifically, the purpose of the study was to (a) assess the separate and combined effects of Self-monitoring Training (SMT) and Reliability Enhancement Package (REP) procedures on the self-monitoring accuracy of speech anxious undergraduates, (b) compare SMT to a procedure designed to control for exposure to irrelevant training stimuli (i.e., Training Control-TC) and REP to a manipulation composed of accuracy instructions and a simple "record" cue (i.e., Accuracy Instructions-AI), and (c) evaluate the effectiveness of experimental procedures when the demand for behavior change was either "low" (unmanipulated) or "high" (manipulated).

Twenty-eight (10 male, 18 female) speech anxious undergraduates, who had indicated an interest in participating in a speech anxiety treatment program, served as subjects. Seven subjects were included in each factorial combination (i.e., SMT-REP, SMT-AI, TC-REP, TC-AI) and all
28 participated in the three self-monitoring speech sessions (i.e., SM I, SM II, and SM III). Participants were initially informed that treatment would consist of repeated exposure (i.e., flooding) to the actual feared situation (i.e., public speaking) and that they would be responsible for evaluating their progress by self-monitoring objective signs of anxiety (i.e., frequency of speech disfluencies) as well as subjective perceptions of fear (i.e., internal sensations of anxiety along an 11-point scale) during each speech performance. Subjects self-monitored objective and subjective behaviors during three separate speech sessions (SM I, II, and III). The three speeches given during each speech session were unobtrusively recorded, which allowed for a comparison to be made between the number of disfluencies self-monitored and the number actually emitted. Experimental procedures were administered between SM I and SM II and demand for behavior change was manipulated prior to the final speech session (i.e., SM III).

The results of the present study revealed that (a) SMT was more effective than TC in increasing the accuracy of self-monitored data, (b) REP as a sole accuracy-enhancing manipulation was more effective than AI only when the target behavior was easily discriminable, and (c) a combination of SMT and REP procedures proved most effective in improving the accuracy of self-monitored data and in maintaining these gains over time. Moreover, subjects receiving the SMT-REP
procedural package were able to eliminate over 50 percent of their self-monitoring errors, double the percentage of speeches monitored "accurately," and increase the correlation between self-recordings and criterion ratings from .56 at pre-manipulation (SM I) to .71 at post-manipulation (SM II). Additionally, results indicated that the speech anxiety treatment program, which served as a medium to study self-monitoring accuracy, was actually quite effective in that subjects showed significant declines in public speaking anxiety and the number of speech disfluencies emitted over the course of the program.

The methodological implications of the current results were discussed in terms of the problems inherent in reinforcement-oriented approaches to the modification of self-monitoring accuracy and the potential for SMT and REP procedures to overcome some of these difficulties. Additional comments were directed toward the deception involved in the cognitive consistency component of the REP procedure and that modification of this manipulation would have to be made before REP would be suitable for use in applied clinical settings.


REFERENCES


Kazdin, A. E. Behavior modification in applied settings. Homewood, Ill.: Dorsey Press, 1972. (a)


Simkins, L. The reliability of self-recorded behavior. Behavior Therapy, 1971, 2, 83-87. (a)


APPENDIX A

PERSONAL REPORT OF CONFIDENCE
AS A SPEAKER (PRCS)
PERSONAL REPORT OF CONFIDENCE AS A SPEAKER (PRCS)

This instrument is composed of 30 items regarding your feelings of confidence as a speaker. After each question there is a "true" and a "false." Try to decide whether "true" or "false" most represents your feelings associated with your most recent speech, then fill in the appropriate "true" or "false" circle. Work quickly and don't spend much time on any one question. We want your first impression on this questionnaire. Now go ahead, work quickly, and remember to answer every question.

1. I look forward to an opportunity to speak in public. (T) (F)
2. My hands tremble when I try to handle objects on the platform. (T) (F)
3. I am in constant fear of forgetting my speech. (T) (F)
4. Audiences seem friendly when I address them. (T) (F)
5. While preparing a speech I am in a constant state of anxiety. (T) (F)
6. At the conclusion of a speech I feel that I have had a pleasant experience. (T) (F)
7. I dislike to use my body and voice expressively. (T) (F)
8. My thoughts become confused and jumbled when I speak before an audience. (T) (F)
9. Although I am nervous just before getting up I soon forget my fears and enjoy the experience. (T) (F)
10. I have no fear of facing an audience. (T) (F)
11. I face the prospect of making a speech with complete confidence. (T) (F)
12. I feel that I am in complete possession of myself while speaking. (T) (F)
13. I prefer to have notes on the platform in case I forget my speech. (T) (F)
14. I like to observe the reactions of my audience to my speech. (T) (F)
15. Although I talk fluently with friends I am at a loss for words on the platform. (T) (F)
16. I feel relaxed and comfortable while speaking. (T) (F)
17. Although I do not enjoy speaking in public I do not particularly dread it. (T) (F)
18. I always avoid speaking in public if possible. (T) (F)
19. The faces of my audience are blurred when I look at them. (T) (F)
20. I feel disgusted with myself after trying to address a group of people. (T) (F)
21. I enjoy preparing a talk. (T) (F)
22. My mind is clear when I face an audience. (T) (F)
23. I am fairly fluent. (T) (F)
24. I perspire and tremble just before getting up to speak. (T) (F)
25. My posture feels strained and unnatural. (T) (F)
26. I am fearful and tense all the while I am speaking before a group of people. (T) (F)
27. I find the prospect of speaking mildly pleasant. (T) (F)
28. It is difficult for me to calmly search my mind for the right words to express my thoughts. (T) (F)
29. I am terrified at the thought of speaking before a group of people. (T) (F)
30. I have a feeling of alertness in facing an audience. (T) (F)
APPENDIX B

DATA SHEET I: OBJECTIVE ANXIETY
DATA SHEET I
OBJECTIVE ANXIETY

SPEECH 1:
[12] [13] [14] [15] [16] [17] [18] [19] [20]
[21] [22] [23] [24]

[12] [13] [14] [15] [16] [17] [18] [19] [20]
[21] [22] [23] [24]

SPEECH 2:
[12] [13] [14] [15] [16] [17] [18] [19] [20]
[21] [22] [23] [24]

[12] [13] [14] [15] [16] [17] [18] [19] [20]
[21] [22] [23] [24]

SPEECH 3:
[12] [13] [14] [15] [16] [17] [18] [19] [20]
[21] [22] [23] [24]

[12] [13] [14] [15] [16] [17] [18] [19] [20]
[21] [22] [23] [24]

119
APPENDIX C

DATA SHEET II: SUBJECTIVE ANXIETY
### DATA SHEET II
### SUBJECTIVE ANXIETY

#### SPEECH 1:

<table>
<thead>
<tr>
<th>No Anxiety Whatsoever</th>
<th>Moderate Anxiety</th>
<th>Panic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### SPEECH 2:

<table>
<thead>
<tr>
<th>No Anxiety Whatsoever</th>
<th>Moderate Anxiety</th>
<th>Panic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### SPEECH 3:

<table>
<thead>
<tr>
<th>No Anxiety Whatsoever</th>
<th>Moderate Anxiety</th>
<th>Panic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D

DATA SHEET SUMMARY
## DATA SHEET SUMMARY

### SESSION I:

<table>
<thead>
<tr>
<th></th>
<th>Speech 1</th>
<th>Speech 2</th>
<th>Speech 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of Repetitions</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Total number of &quot;Ah's&quot;</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Anxiety Level</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
</tbody>
</table>

### SESSION II:

<table>
<thead>
<tr>
<th></th>
<th>Speech 1</th>
<th>Speech 2</th>
<th>Speech 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of Repetitions</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Total number of &quot;Ah's&quot;</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Anxiety Level</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
</tbody>
</table>

### SESSION III:

<table>
<thead>
<tr>
<th></th>
<th>Speech 1</th>
<th>Speech 2</th>
<th>Speech 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of Repetitions</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Total number of &quot;Ah's&quot;</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Anxiety Level</td>
<td>___</td>
<td>___</td>
<td>___</td>
</tr>
</tbody>
</table>

123
APPENDIX E

SPEECH TOPICS AND SUBTOPICS
SPEECH TOPICS AND SUBTOPICS

I. Session 1 (SM I)

A. What are your opinions regarding the Women's Liberation Movement?
   1. Job and wage equality.
   2. Women in government.
   3. Women in the armed forces.
   4. Women as homemakers.

B. What are your opinions regarding sororities and fraternities?
   1. Potential benefits for members.
   2. Negative aspects.
   3. Would you join one?
   4. Do they foster individual growth?

C. Is there life on other planets?
   1. Is it possible?
   2. Unidentified flying objects (UFO's).
   3. Advanced or primitive life.
   4. What are the implications.

II. Session 3 (SM II)

A. Should all forms of gambling be legalized in Montana?
   1. Another Las Vegas.
   3. Increased state revenue.
   4. Increased crime.
B. Should all women have a legal right to abortion?
   1. Your opinion regarding ethics.
   2. In cases of rape.
   3. Negative aspects of illegitimate abortions (by unqualified individuals).
   4. In cases of probable birth defects.

C. Should marijuana be legalized?
   1. Increased government revenue from marijuana tax.
   2. If alcohol is legal, why not pot?
   3. Progression to harder drugs.

III. Session 4 (SM III)

A. Should we continue to fund athletic programs at the University of Montana?
   1. More money for educational purposes.
   2. Losing football teams.
   3. Loss of an entertainment area.
   4. Physical activities are as important as academic pursuits.

B. Should the government appropriate more money for national defense?
   1. Too many arms already.
   2. Possibility of an accidental nuclear holocaust.
   3. Keeping the balance of power.
C. What are your opinions regarding Gerald Ford's performance as President of the United States?

1. Persistent use of the veto.
2. Economy and unemployment.
3. Foreign policy and détente.
4. Will he win the next election?
APPENDIX F

SMT SPEECH TRANSCRIPTS
Speech 1

Segment 1
The United States, um, the United States is uh now talking about do - - doing ah business with - with Fidel Castro, even about re - restoring normal diplomatic relations with Cuba.

"Ah" = 3
Repetitions = 4

Segment 2
But ah noth - nothing is either normal or ah diplomatic in - in Havana these days. Ah, the population, the population of the ah island has increased from - from six um six million to ah nine million in the 15 years since, - since ah President Eisenhower broke relations.

"Ah" = 7
Repetitions = 6

Segment 3
About half the - the present um population is now, ah, is now under 18 and - and has no mem-memory of any political system but Cuban communism. It's a different - a different society and a ah different ah generation, born into the um cold - cold war, educated in ah hostility toward the ah United - United
States, and emerging into the dubious world of detente. Uh, even Castro, now in his 50th year, is um beginning to whistle a different tune. My husband and I spent four 18-hour days with Fidel Castro last August at the uh invitation of President Echeverria of Mexico, who was visiting um Castro for the first time.

"Ah" = 11
Repetitions = 12

Segment 4
We talked privately for ah more than an hour about ah U.S.-Cuba re-relations, and Castro impressed me as a man who had made a revolution in his youth and ah now, in middle age, was confronted by the ah more tedious task of governing a country in a disorderly, changing world. His fears of invasion and ah defeat have have passed, but um so has the exhilaration of the ah struggle. Ah, he has more composure and ah dignity, and seems more disciplined, both um mentally and phy physically, than when I saw him six years ago.

Ideologically and ah economically, Castro is is still tied to the um communist system. He still har-harbors in Havana, and finances the left-wing leaders of a small Puerto Rican independence movement, but ah his revolutionary movement in the rest of Latin America has collapsed. Now he talks more about um importing food, machinery
and ah modern technology than um about exporting his - _his_ revolutionary ideas to them.

"Ah" = 16
Repetitions = 18

Speech 2

Segment 1
Great Britain is - _is_ sick. Everywhere you look the um evidence abounds. The out - _outward_ signs are a pound sterling that is not - _not_ merely declining but ah shrinking.

"Ah" = 2
Repetitions = 3

Segment 2
The inward signs are - _are_ no less um evident. With few exceptions, the - _the_ physical plant of Great - _Great_ Britain's industries is decrepit; its steel mills and ah automobile factories are trying to - _to_ make do with ah outmoded and um worn - _worn-out_ machinery. Few of its industries can uh compete with its partners in the Common Market or uh with the outside world.

"Ah" = 6
Repetitions = 5
Segment 3
With all this, the standard of living of its people is um lower than that of comparable countries in uh Europe; vastly lower than that of the United States. Uh, and it is a shrinking standard of living for all, peer and ah plowman alike. Time is running out even on those in ah protected, subsidized industries. It has already run out on the um middle class doctors, accountants, college professors, clerks, um journalists, civil servants, and um shop keepers. It is all very curious. For Britain has not been ah brought to this es estate by um defeat in war or by any natural disasters.

"Ah" = 10
Repetitions = 11

Segment 4
Britain's undoing is its own doing. It has been brought to this um largely by the policies of its government, and uh by the resigned acceptance of the people. Thus, Britain offers a model study in how to um ruin a once vigorous nation.
The formula is uh simple. You begin by putting upon a nation an economic burden it can - cannot bear. In ah Britain's case, it was an all-encompassing welfare program; including a uh free medical program, uh subsidized housing, subsidized food, and an subsidized transportation.
One way or another, all this - all this must be paid for. Ah this means either higher taxes or a um resort to the government printing presses to - to create money - or both. The ah government-printed money causes - causes inflation, which in - increases the ah cost of - of living, including ah the welfare program, which in turn - which in turn calls for more ah printed money, accelerating the ah inflation.

"Ah" = 15

Repetitions = 17
APPENDIX G

SPEECH RATING FORM
# SPEECH RATING FORM

## SPEECH I:

<table>
<thead>
<tr>
<th></th>
<th>Extremely Fluent</th>
<th>Moderately Fluent</th>
<th>Extremely Disfluent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Not At All Anxious</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Well Organized</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

## SPEECH II:

<table>
<thead>
<tr>
<th></th>
<th>Extremely Fluent</th>
<th>Moderately Fluent</th>
<th>Extremely Disfluent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Not At All Anxious</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Well Organized</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
APPENDIX H

PROGRAM EVALUATION QUESTIONNAIRE
PROGRAM EVALUATION QUESTIONNAIRE

Directions: Please read each of the following questions carefully before responding. Answer each question in the order in which it is presented and do not go back and change any of your responses once you have committed yourself.

1. Would you have participated in the speech anxiety treatment program if you had not been offered experimental credit for participation? (check one)

_________(a) yes  _________(b) not sure  _______(c) no

2. In your own words, what was it that was being investigated (i.e., studied) during the current speech anxiety program?

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________

3. Was there anything about the present program that you did not understand or that aroused your suspicion? (please specify)

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
4. **Before** coming to today's session, how much of a decline in **subjective** anxiety did you expect to experience while giving your final three speeches? (check one)

- (a) A dramatic decline
- (b) A sizable decline
- (c) A moderate decline
- (d) A slight decline
- (e) No decline whatsoever

5. **Before** coming to today's session, how much of a decline in **objective** anxiety (i.e., number of "Ah's" and Repetitions) did you expect to find while giving your final three speeches? (check one)

- (a) A dramatic decline
- (b) A sizable decline
- (c) A moderate decline
- (d) A slight decline
- (e) No decline whatsoever

6. In your opinion, how accurate were you in monitoring and recording "Ah's" and Repetitions?

<table>
<thead>
<tr>
<th>Extremely Inaccurate</th>
<th>Moderately Accurate</th>
<th>Extremely Accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. How hard did you **try** to accurately self-monitor "Ah's" and Repetitions?

<table>
<thead>
<tr>
<th>Didn't Try At All</th>
<th>Put a Moderate Amount of Effort Into It</th>
<th>Put Out Maximal Effort to Catch Every Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>