EFFECTS OF SELF-SCORING ON TEACHERS’ RATES OF POSITIVE
AND NEGATIVE STATEMENTS DURING CLASSROOM INSTRUCTION

DISSERTATION

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

By
Susan M. Silvestri, M. A.

* * * * *

The Ohio State University
2004

Dissertation Committee:
Professor William L. Heward, Adviser
Professor James H. Cowardin
Professor Stephanie M. Peterson

Approved by:

Adviser
College of Education
Teacher attention, in the form of positive statements made to students, is one of the most important variables in the design and delivery of effective instruction. Many teachers, however, reserve their attention for times when students misbehave, instead of using their attention and approval to encourage appropriate student behavior.

Although studies have demonstrated the effectiveness of teacher praise in improving student behavior, there is little research on strategies to increase teachers’ positive statements. One proposed strategy is training teachers to score brief audiotapes of their instruction for the occurrence of positive statements, graphing that number, and setting a goal for subsequent instructional sessions.

This study examined the effects of self-scoring on teachers’ subsequent positive statements during classroom instruction. Four teachers employed at a large Midwestern elementary charter school participated. Instructional sessions were recorded daily and generalization sessions were recorded weekly.

The dependent variables of interest included the rates of generic positive statements, behavior-specific positive statements, and negative statements, and the percentage of statements repeated within each session. The self-scoring intervention consisted of meetings with the experimenter in which participating teachers selected a 5-minute sample of an audiotaped session, timed and recorded their positive statements, graphed the total number of positive statements, and set a goal for the total number of
positive statements for the next recording session. Meetings took place four times per week during the first week of intervention, three times during the following week, and so on, until they were occurring once per week. One participant took part in two additional conditions that assessed the effects of self-scoring generalization videotapes, and one participating teacher took part in a condition in which she selected an additional tactic (using a tactile cueing device) to aide in increasing positive statements after self-scoring alone was demonstrated to be ineffective.

Results demonstrated increases in positive statements for three of four participants. The rate of negative statements decreased for all four participants. Three participants showed moderate, but short-lived increases in positive statements during generalization sessions. Results are discussed in terms of the function of the self-scoring sessions. Limitations of the study are addressed.
ACKNOWLEDGMENTS

I wish to sincerely thank Dr. William Heward for the guidance and support he has provided during my dissertation work and throughout my doctoral program. It is thanks to him that my doctoral program exceeded my expectations.

I would also like to thank my dissertation committee members, Dr. Stephanie Peterson and Dr. James Cowardin, for their encouragement and input over the last three years.

Thank you to Sara Ernsbarger and Terri Hessler for their assistance with social validity assessments, and to Erin Gady, Nicole Hinsey, and Sarah Painter for their help with interobserver agreement data collection. I sincerely appreciate their contributions; I couldn’t have completed my dissertation without their help.

I owe a special thank you to Barry Merrick and Phil Fahy, who provide on-site technical assistance, and also to the faculty, staff, and students at the school who so generously hosted this study.

Finally, I am indebted to my participants, who welcomed me into their classrooms. I appreciate their flexibility and their willingness to be a part of my study. I hope they learned as much from me as I learned from them.
VITA

March 16, 1973 .............................................................. Born – New Haven, CT

1995.............................................................................. B.A. Psychology, Smith College

1995-1998 ......................................................... Teacher, New England Center for Children
Southboro, MA

1998-2001 ............................................................ Teacher, Fred S. Keller School
Yonkers, NY

2000.......................................................... M.A. Behavior Disorders, Columbia University Teachers College

2002-present................................................ Graduate Assistant, The Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Education
Special Education
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td><strong>Chapters:</strong></td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>3</td>
</tr>
<tr>
<td>2. Review of the Literature</td>
<td>4</td>
</tr>
<tr>
<td>Effects of Praise and Approval on Student Behavior</td>
<td>5</td>
</tr>
<tr>
<td>Descriptive Studies on Teacher Praise</td>
<td>12</td>
</tr>
<tr>
<td>Strategies to Increase Teachers’ Use of Positive Statements</td>
<td>17</td>
</tr>
<tr>
<td>Cueing Systems</td>
<td>17</td>
</tr>
<tr>
<td>Public Posting</td>
<td>19</td>
</tr>
<tr>
<td>Instructions and Feedback</td>
<td>20</td>
</tr>
<tr>
<td>Student Recruiting</td>
<td>24</td>
</tr>
<tr>
<td>Goal Setting</td>
<td>29</td>
</tr>
<tr>
<td>Self-Scoring and Self-Evaluation</td>
<td>31</td>
</tr>
<tr>
<td>Individually Determined Interventions</td>
<td>38</td>
</tr>
<tr>
<td>Summary of Literature</td>
<td>38</td>
</tr>
<tr>
<td>Research Questions</td>
<td>39</td>
</tr>
<tr>
<td>3. Method</td>
<td>41</td>
</tr>
<tr>
<td>Participants</td>
<td>41</td>
</tr>
<tr>
<td>Teacher 1</td>
<td>42</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>43</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>44</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>44</td>
</tr>
<tr>
<td>Setting</td>
<td>45</td>
</tr>
<tr>
<td>Experimenter</td>
<td>46</td>
</tr>
<tr>
<td>Data Collectors</td>
<td>47</td>
</tr>
<tr>
<td>Materials</td>
<td>47</td>
</tr>
</tbody>
</table>
Definition and Measurement of Dependent Variables ...............................................47
Definitions..............................................................................................................47
Generic Positive Statements ..............................................................................48
Behavior-Specific Positive Statements ............................................................. 48
Negative Statements.......................................................................................... 49
Repeated Statements.......................................................................................... 50
Measurement and Recording .................................................................................51
Procedures to Enhance Believability of Data.............................................................52
Interobserver Agreement .......................................................................................52
Treatment Integrity ................................................................................................53
Social Validity .......................................................................................................54
Experimental Design ..................................................................................................55
Procedures ..................................................................................................................56
Baseline..................................................................................................................56
Training ..................................................................................................................56
Self-Scoring ............................................................................................................57
Self-Scoring plus MotivAider® .............................................................................59
Self-Scoring of Generalization Tapes ............................................................... 60
Self-Scoring of Generalization Tapes; No Microphone .................................... 60
Videotape Probes ...................................................................................................61
4. Results ..................................................................................................................... ...62
Interobserver Agreement ............................................................................................62
Treatment Integrity .....................................................................................................63
Positive and Negative Statements ..............................................................................64
Teacher 1................................................................................................................64
Baseline .............................................................................................................64
Self-Scoring ............................................................................................................66
Self-Scoring of Generalization Tapes ............................................................... 66
Self-Scoring of Generalization Tapes; No Microphone .................................... 68
Repeated Positive Statements............................................................................ 68
Generalization Data........................................................................................... 68
Teacher 2................................................................................................................70
Baseline .............................................................................................................70
Self-Scoring ............................................................................................................71
Repeated Positive Statements............................................................................ 71
Generalization Data........................................................................................... 71
Teacher 3................................................................................................................73
Baseline .............................................................................................................73
Self-Scoring ............................................................................................................73
Self-Scoring plus MotivAider® .............................................................................74
Repeated Positive Statements............................................................................ 74
Generalization Data........................................................................................... 75

vii
C: Sample Self-Scoring Graph

D: Sample Transcripts of Self-Scoring Meetings

E: Menu of Intervention Choices

F: Picture of MotivAider®

G: Treatment Acceptability Rating Form (Modified)

H: Social Validity Survey
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>IOA data for participants across dependent variables</td>
<td>63</td>
</tr>
<tr>
<td>4.2</td>
<td>Means and ranges of dependent variables by Teacher 1</td>
<td>67</td>
</tr>
<tr>
<td>4.3</td>
<td>Means and ranges of dependent variables by Teacher 2</td>
<td>70</td>
</tr>
<tr>
<td>4.4</td>
<td>Means and ranges of dependent variables by Teacher 3</td>
<td>75</td>
</tr>
<tr>
<td>4.5</td>
<td>Means and ranges of dependent variables by Teacher 4</td>
<td>77</td>
</tr>
<tr>
<td>4.6</td>
<td>Treatment acceptability rating scores by participants</td>
<td>84</td>
</tr>
<tr>
<td>4.7</td>
<td>Social validity survey responses by current and future teachers</td>
<td>85</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Positive and negative statements per minute by Teacher 1</td>
<td>65</td>
</tr>
<tr>
<td>4.2</td>
<td>Positive and negative statements per minute by Teacher 2</td>
<td>69</td>
</tr>
<tr>
<td>4.3</td>
<td>Positive and negative statements per minute by Teacher 3</td>
<td>72</td>
</tr>
<tr>
<td>4.4</td>
<td>Positive and negative statements per minute by Teacher 4</td>
<td>76</td>
</tr>
<tr>
<td>4.5</td>
<td>Percentage of positive statements repeated by all participants</td>
<td>79</td>
</tr>
<tr>
<td>4.6</td>
<td>Positive statements per minute by all participants</td>
<td>81</td>
</tr>
<tr>
<td>4.7</td>
<td>Negative statements per minute by all participants</td>
<td>82</td>
</tr>
<tr>
<td>4.8</td>
<td>Mean rate of statements by all participants across conditions</td>
<td>83</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Teacher praise is an important tool for changing students’ behavior. Since the early 1960s many experimental studies have demonstrated that teacher praise can function as a positive reinforcer, producing increases in student behaviors such as academic responding and on-task behavior. In addition, increasing teacher praise has been demonstrated to be an effective tactic across a broad range of educational settings, including elementary, secondary, and special education classrooms (Craft, Alber, & Heward, 1998; Sutherland, 2000).

Early and ongoing research in applied behavior analysis and special education demonstrated the effectiveness of teacher praise, attention, and approval in improving a variety of student behaviors (Madsen, Becker, & Thomas, 1968). Praise and positive attention from a teacher can function as a reinforcer for study behavior (Hall, Lund, & Jackson, 1968), for quiet on-task behavior (Thomas, Becker, & Armstrong, 1968; van der Mars, 1989; Ferguson & Houghton, 1992; Sutherland, Wehby, & Copeland, 2000), and for academic behaviors such as writing digits in the correct order, even in the absence of explicit instruction or correction (Hasazi & Hasazi, 1972).
Several observational studies have found that many teachers do not provide students with sufficient rates of praise to maintain appropriate behavior and correct responses in the classroom (White, 1975; Ysseldyke, Thurlow, Mecklenburg, & Graden 1984; Wyatt & Hawkins, 1987; Wheldall, Houghton, & Merrett, 1989; Harrop & Swinson, 2000). Too often, teachers reserve their attention for use as a punitive consequence when students behave inappropriately.

Many experts in regular and special education agree that teachers should praise their students more often, yet few empirical studies have demonstrated effective ways to train teachers to do this (Sutherland, 2000). Potential strategies for increasing teachers’ rates of praise include cueing systems (Van Houten & Sullivan, 1975), public posting (Gross & Ekstrand, 1983), instructions and feedback (Cooper, Thomson, & Baer, 1970), teaching students to recruit positive teacher attention (Craft, Alber, & Heward, 1998), goal setting (Martens, Hiralall, & Bradley, 1997), and self-scoring (Sutherland & Wehby, 2001), however little published research exists on these topics, so it is not surprising that teachers generally do not use praise as a classroom management technique.

Self-scoring is a promising technique for training teachers to increase their use of positive statements during instruction. This strategy involves teachers scoring audiotapes of their own instruction for the occurrence of positive statements, graphing the number of positive statements recorded, setting goals for the next observation, and praising themselves for their progress (Sutherland & Wehby, 2001). Previous investigations of self-scoring have included minimal contact with an experimenter to assist teachers with
scoring, goal setting, and praising, and have employed group, instead of single-subject, designs.

Purpose of the Study

The purpose of this study was to assess the effects of self-scoring audiotapes of instruction on teachers’ subsequent rates of positive statements and negative statements, and the effects of an intervention package consisting of self-scoring audio and videotapes of instruction and individually determined interventions on teachers’ subsequent rates of positive statements and negative statements. Teachers met with the experimenter to score audiotapes of their instruction and record the number of positive statements made during a 5-minute segment of a taped teaching session. The teachers then graphed the number of positive statements they had recorded and set a goal for their next self-scoring session. The experimenter encouraged teachers to continue to increase their positive statements, and praised teachers when they met their goals. The teacher and experimenter also discussed possible strategies that the teacher could use to increase positive statements in future lessons.
CHAPTER 2

REVIEW OF THE LITERATURE

Teacher-mediated positive reinforcement is an important part of effective classroom instruction. Both positive and negative reinforcement operate in classrooms. Positive reinforcement most often occurs in the form of attention from the teacher, and results in increases in the behaviors it follows. Positive reinforcement occurs when students are told that their answers are correct, and when they are recognized and rewarded for behaving in accordance with classroom rules (and continue to behave appropriately in the future). When teachers utilize positive reinforcement to change behavior, students are eager to learn and to interact with their teachers. Negative reinforcement also produces increases in student behavior; however negative reinforcement produces behavior that avoids attention from the teacher. For example, students may remain on task and provide correct responses not because of doing so has produced praise and recognition in the past, but rather because not doing so has previously produced negative consequences. Negative reinforcement is effective in changing student behavior, but it also conditions the teacher and the classroom as stimuli to be avoided.

Some teachers take the time to develop elaborate systems of reinforcement, including tokens or other generalized reinforcers, reinforcer menus and pricing systems,
and multiple levels of back-up reinforcers. The simplest reinforcer for teachers to deliver is praise in the form of positive statements about student behavior. Positive statements and social attention are important elements of social and academic learning. Positive statements require little effort to deliver, can be used without any special materials, and can function as reinforcers; positive statements can and should be used frequently by teachers when students are demonstrating desirable social and academic behavior in order to maintain that behavior.

This chapter will review the literature on the effects of praise on student behavior, along with several descriptive studies documenting natural approval rates in classrooms. Next, strategies to increase teachers’ use of praise and positive feedback will be reviewed; these strategies include cueing systems, public posting, instructions and feedback, training students to recruit attention, goal setting, self-scoring and self evaluation, and the effects of having teachers self-select the interventions they would like to use. Last, the research questions for this study will be presented.

**Effects of Praise and Approval on Student Behavior**

Throughout the history of applied behavior analysis research has demonstrated the effectiveness of praise in the improvement of social and academic behavior. Zimmerman and Zimmerman (1962) altered the contingencies that were maintaining the disruptive behavior of two 11-year-old boys with emotional disturbance. Classroom observations demonstrated that both boys frequently received individual attention, often from more than one adult, contingent on disruptive and off-task behavior. The first author, who served as the student’s English instructor, manipulated the consequences for disruptive
behavior both by ignoring off-task behavior and by providing differential social consequences in the form of vocal and non-vocal approval and physical proximity. Both boys demonstrated decreases in disruptive target behaviors and increases in appropriate behavior and correct academic responses.

Another investigation that manipulated teacher attention was conducted by Hall, Lund, and Jackson (1968). Six students participated in this study, one from a first grade class and five from a third grade class. During baseline sessions student study behavior, defined as “orientation toward the appropriate object or person,” (p. 2) was observed and recorded. During intervention phases the experimenter held up a small piece of colored paper to cue the teacher to provide praise and attention to a particular student. Results demonstrated that student study behavior increased when the cueing system was implemented and also that study behavior was maintained even after the cueing system was discontinued.

In 1968 Thomas, Becker, and Armstrong published a powerful demonstration of the effects of two kinds of teacher behavior: approval and disapproval. Approving behavior consisted of physical contact, verbal, and facial subclasses. Physical contact included behaviors such as embracing or patting students; verbal behavior included praise and approval; and facial behavior included smiling, winking, and nodding. Disapproving behavior also consisted of physical contact, verbal, and facial subclasses. Physical contact included holding, hitting, and shaking students; verbal behavior included yelling, scolding, threatening, and belittling; and facial behavior included frowning, grimacing, and head shaking. Student behaviors measured included disruptive behavior and relevant
(appropriate) behavior. The study was conducted with teachers who were observed to emit high rates of approval before intervention began. Thus, baseline sessions were associated with high rates of approving behavior from the teacher. The study was designed to demonstrate the effects of teacher disapproval on student disruptive behavior.

During the first baseline phase no attempt was made to change the teachers’ behavior. In the second phase the teachers demonstrated moderate rates of disapproval and low rates of approval (called “no approval”); this phase was followed by a return to baseline conditions. In the fourth phase the moderate disapproval/low approval conditions were implemented. This was followed by a “frequent disapproval” condition, a “no approval” phase, and a return to baseline.

Student disruptive behavior was consistently higher in the “no approval” and “frequent disapproval” phases, and lowest in the baseline phases. In addition, relevant appropriate behaviors were highest in the baseline phases. This is an important study because it demonstrates not only that praise and approval can produce high rates of on-task and low rates of disruptive behavior, but also that high rates of disapproval are often related to low rates of appropriate and high rates of disruptive behavior. The authors made the important point that overall, teacher attention functioned as a reinforcer for the behavior of these students. The nature of that attention (approving or disapproving) was not relevant. If students could only acquire teacher attention by emitting disruptive behavior, then they would emit disruptive behavior. “For some children any adult attention may be reinforcing” (p. 44). Thomas and colleagues concluded with an important warning for teachers: “unless an effort is made to support desirable classroom
behaviors with appropriate consequences, the children’s behavior will be controlled by others in ways likely to interfere with the teacher’s objectives” (p. 45).

In another study that demonstrated the effects of teacher approval, Madsen, Becker, and Thomas (1968) assessed the effectiveness of three elements of classroom behavior management: rules, praise, and ignoring, with the behaviors of three elementary school students. Their data demonstrated that rules alone were not sufficient for improving student behavior. Ignoring inappropriate behavior and praising appropriate classroom behavior were much more effective in increasing desired behavior in the classroom. They concluded that the essential component of their intervention, and of any other effective classroom management package, is praise for appropriate behavior.

In a study by van der Mars (1989) the effects of a teacher’s specific verbal praise on student off-task behavior were examined in the context of a physical education class. Participants included two boys and one girl from a second-grade class. A multiple baseline design was employed to examine the effects of increased teacher praise on off-task behavior. In this study, audio cues presented via a micro-cassette player were used to remind the teacher to praise the students. Results demonstrated a decrease in off-task behavior when rates of praise increased. In addition, the audio cueing system was demonstrated to be effective in prompting the teacher to praise. The author concluded, “Physical educators predominantly correct, nag, or desist students to reduce inappropriate conduct. Proper training in the contingent use of specific, genuine praise could help change this behavior pattern” (p. 168).
Hasazi and Hasazi (1972) manipulated teacher attention and examined the effects on the digit-reversal behavior of an 8-year-old boy. The student was identified by the teacher as a “capable” (p. 158) student, however he demonstrated difficulty in adding numbers that produced a two-digit sum. He would reverse the order of the two digits in most problems. Baseline sessions consisted of the teacher marking all correct answers with a “C” and all digit reversals with an “X”. The teacher then commented on the digit reversals. In the intervention phase all digit reversals and all correct responses were marked with a “C”. In addition, correctly written sums were followed by differential approving consequences from the teacher, including praise, a smile, and a pat on the back. These two phases were followed by a return to baseline and a replication of the intervention. Results demonstrated dramatic reductions in digit reversals during intervention phases, and high rates of digit reversals during baseline phases, clearly indicating that digit reversals were under the control of teacher attention. No specific math instruction was required to reduce digit reversals. The re-allocation of teacher attention to correctly written sums sufficed to reinforce correct responding.

Two studies by Kazdin also demonstrate the effectiveness of teacher attention in changing student behavior. Kazdin (1973) and Kazdin and Klock (1973) examined the relationship between teacher approval and student attentive behavior by systematically manipulating teacher approval. Kazdin (1973) collected baseline rates on student attentive behavior with two pairs of students (one target and one non-target student in each pair), and then assessed the attentive behavior of both the target and the non-target student as praise was delivered to the target student. Both target students and both non-
target students showed increases in attentive behavior, demonstrating that praise was
effective in improving behavior not only directly for the student whose behavior was
praised, but also vicariously for a neighboring student.

Kazdin and Klock (1973) also demonstrated the effectiveness of non-vocal approval
in improving the attentiveness of 12 students with moderate disabilities. An ABAB
design was employed to examine the effects of increasing the teacher’s non-vocal
approval (e.g., smiling, physical contact) on student attentive behavior. Attentive
behavior was defined as sitting in the assigned seat, working on the given assignment,
and not talking without permission. During intervention phases the teacher was instructed
to increase her non-vocal approval, but not her vocal (i.e., praise) approval behaviors.
Results demonstrated consistent increases in student attentive behavior when the teacher
provided higher rates of non-vocal approval.

Ferguson and Haughton (1992) examined the effects of increased rates of teacher
praise on students’ on-task behavior. Three teachers, each from a different elementary
school, volunteered to participate. Eight target students in each class were observed in a
rotating fashion for 5 seconds each. Observers recorded student on-task behavior with a
whole-interval recording system for three minutes, and then counted teachers’ positive
and negative statements for three minutes; each observation lasted for 30 minutes.

After baseline data were collected (4 observations in teacher A’s classroom, 7
observations in teacher B’s classroom, and 10 observations in teacher C’s classroom) the
teachers were instructed in the appropriate use of contingent positive statements and were
asked to deliver at least one contingent positive statement to each of the target students
during each 30-minute observation. Results demonstrated that in classroom A, the teacher’s rate of praise statements increased by 36%, and target students’ on-task behavior increased 12%. In classroom B, the teacher’s praise increased by 127%, and target students’ on-task behavior increased by 20%. In classroom C, the teacher’s praise increased by 85%, and target students’ on-task behavior increased by 12%.

Sutherland, Wehby, and Copeland (2000) examined the effects of varying rates of praise on students’ on-task behavior. The teacher participant was a male teacher of students with emotional and behavioral disorders (EBD) with three years of teaching experience. The student participants were two girls and seven boys, with ages ranging from 10-11. All sessions were conducted in a fifth-grade classroom for students identified with EBD. Dependent variables for the teacher included non-behavior-specific praise (measured with a frequency count) and behavior-specific praise (measured with a frequency count). Student on-task behavior was measured with a momentary time-sampling procedure. For the purposes of data collection for student behavior, the classroom was divided into four areas; these were observed for 1-minute intervals on a rotating basis. In order for on-task behavior to be recorded for a given area, all students in that area had to demonstrate on-task behavior for the recording period. Praise data were reported as the total number of statements observed per session, and on-task behavior data were converted to a percentage of total intervals observed.

An ABAB reversal design was used to demonstrate the effects of increased praise on student on-task behavior. Baseline data demonstrated moderate levels of student on-task behavior (mean: 48.7%) and moderate numbers of behavior-specific (mean: 1.3) and non-
behavior-specific (mean: 3.3) praise. The intervention consisted of instructing the teacher on the importance of using behavior-specific praise. The observer (one of the authors) gave the teacher examples of behavior-specific praise, and also gave him feedback on his observed rate of behavior-specific praise. The teacher and the observer agreed upon a goal of six behavior-specific praise statements during each observation. Prior to each observation, the observer reminded the teacher of his goal. Immediately after each observation the observer gave the teacher feedback on his performance.

During the intervention phase, behavior-specific praise statements increased to a mean of 6.7 and non-behavior-specific praise increased to a mean of 3.7. Student on-task behavior increased to 85.6% in this phase. Both types of positive statements and student on-task behavior decreased when the intervention was removed, and all dependent variables increased (behavior-specific praise mean: 7.8; non-behavior specific praise mean: 4.7; on-task behavior: 83.3% of observed intervals) when the intervention was reinstated.

These studies demonstrate the effectiveness of praise and approval as consequences for appropriate behavior, and also the powerful effects of teacher attention in general, regardless of whether it is positive or punitive in nature. Teachers can use their positive attention and approval to improve a variety of student behaviors in a wide range of settings.

Descriptive Studies on Teacher Praise

Although the positive effects of teacher praise on student behavior and academic performance have been demonstrated in empirical literature, few teachers systematically
employ rates of praise that are sufficient to maintain students’ on-task and appropriate behavior.

White (1975) conducted one of the first studies to observe teachers’ use of praise across a large number of teachers (104) and a wide range of grade levels (1-12). White summarized data collected during 16 systematic observations. In all, the total observation time was 8340 minutes; data were converted to number per minute. White collected data on teacher approval and disapproval. The results showed higher rates of approval in earlier grades, with a sharp decline in approval after the second grade. The rate of approval continued to decline through high school where it stabilized at about one approval statement every 5 to 10 minutes. Disapproval behaviors showed a similar pattern, peaking in early grades and declining with older students. White also noted an important trend regarding the ratio of approval to disapproval: teachers in earlier grades tended to provide more approval than disapproval; however, starting in the second grade disapproval became more frequent than approval.

White also divided approval and disapproval into instructional and managerial categories. “Instructional” referred to student behavior related to the on-going activity in the class, whereas “managerial” referred to classroom management, including talking out and other misbehavior. Teacher approval rates for instructional behavior were higher than approval rates for managerial behavior across grades; rates of disapproval for managerial behavior were higher than approval rates for the same category. In addition, teachers in grades five and higher used more disapproval for managerial behavior than approval for instructional behavior. For the most part, approval for managerial behavior was virtually
nonexistent, even in primary grades; beginning in the fourth grade, students were more likely to receive disapproval for managerial behavior.

White (1975) discussed possible reasons for the low rates of teacher approval behaviors she found. She made the important point that the use of disapproval is more reinforcing for teachers. When a teacher reprimands a student who is misbehaving, the student’s behavior usually changes immediately. When a teacher praises an on-task student who is working quietly, the student’s behavior does not immediately change. Over time, the student will emit the behavior more often if praise continues to be delivered, however there is no immediate change in the environment that is discernible to the teacher. “Giving approval for appropriate pupil instructional behaviors may not be particularly reinforcing for teachers, because the giving of approval does not result in any outcome that is obviously due to the immediate effectiveness of the teacher” (p. 371). As White pointed out, when students are learning and on-task, they are just doing what they are supposed to do. As such, teachers may find disapproval more reinforcing than ignoring inappropriate behavior and praising alternative behavior.

Wyatt and Hawkins (1987) conducted a similar analysis of teachers’ use of verbal approval and disapproval. They observed 35 public school teachers in kindergarten, 1st, 2nd, 3rd, 4th, 9th, and 12th grades. Teachers worked at schools located in metropolitan and rural areas. Wyatt and Hawkins used the same observation instrument used by White (1975): the Teacher Approval and Disapproval (TAD) code (White, Beecher, Heller, & Watters, 1973).
In contrast to White’s (1975) findings, Wyatt and Hopkins (1987) did not observe a predominance of disapproval emerge in later grades: in all grades observed, rates of approval were higher than rates of disapproval, although both approval and disapproval declined in higher grades. Like White, they found relatively low rates of approval for conduct, or “managerial” behavior, and higher rates of disapproval for conduct than for instructional behavior. Wyatt and Hawkins emphasized that teachers should be trained in the use of verbal approval.

Wheldall, Houghton, and Merrett (1989) conducted a large-scale observational study on the approval and disapproval behaviors of 130 British secondary school teachers. The participants included 63 female and 67 male teachers across academic subjects. Students ranged in age from 11 to 16. Class size ranged from 8 to 32, with a mean of 22 students. Each observation lasted for 30 minutes and included rotating 3-minute intervals in which observers recorded the frequency of teacher approval and disapproval statements, and then student on-task behavior.

Wheldall et al. (1989) found that the teachers they observed did use slightly more approval than disapproval overall, but similar to previous findings, disapproval for social behavior was three times higher than approval for social behavior. Also, more attention overall was given for academic than for social behaviors. About a quarter of Wheldall et al.’s teachers used twice as many approval as disapproval responses for academic and social behavior combined, but almost three quarters of teachers gave twice as many disapproval as approval responses to social behavior. Analysis of approval and disapproval statements by age group taught shows that approval for social and academic
behavior decreased as students got older. Disapproval for social behavior was variable, but overall was lower than either category of approval behavior. The authors wrote that “It appears that secondary teachers are quick to ‘pick up’ incidents of which they disapprove, yet they rarely approve or comment upon desirable social behaviors” (p. 45).

Another descriptive study of rates of teacher approval and disapproval in British schools was conducted by Harrop and Swinson (2000). These authors summarized observations of 30 teachers. Ten teachers taught in infant school (preschool - early elementary) classrooms, 10 taught in junior school (early –older elementary) classrooms, and 10 taught in secondary school (middle school) classrooms. One lesson per teacher was recorded via a wireless microphone worn by the teacher.

Data confirmed the results of previous studies that had shown higher rates of approval for academic behavior than for social behavior, and higher rates of disapproval for social behavior than academic behavior. Teacher approval for academic behavior was the most frequently recorded teacher behavior, followed by disapproval for social behavior, then disapproval for academic behavior, and finally, approval for social behavior. Approval for social behavior was virtually nonexistent (range across grade levels: .06 - .01 per minute).

Overall, teachers used more approval than disapproval, provided almost no approval for social behavior, and allocated most of their feedback to approval of academic behavior. In addition, disapproval rates decreased as student grade level increased, however data from this study showed that all teacher attentive (approval and disapproval) decreased with student grade level.
In summary, descriptive studies on teachers’ use of positive statements have found that teachers generally do not provide high rates of positive attention for their students’ managerial or social behavior. Teachers also tend to provide less positive feedback as students get older, and give more negative feedback than positive feedback for managerial behavior. These studies illustrate the need to train teachers to increase their positive feedback, in particular the positive feedback they provide for appropriate social behavior.

Strategies to Increase Teachers’ Use of Positive Statements

Multiple interventions have been tested to ameliorate this deficit in teacher skills. Examples include audio cueing, public posting of praise rates, instruction plus feedback, teaching students to recruit positive teacher attention, goal setting, and self-scoring.

Cueing Systems

Cueing systems are prompt strategies that alert a teacher to provide positive attention and feedback to students. Cueing systems can include audio cues (for example, chimes or beeps played on an audiotape), tactile cues (wearing a device that vibrates at predetermined fixed or variable intervals), or visual cues (for example, a person in the back of the classroom holding up a colored signal card), or a combination of these, to prompt the teacher to praise a student or a group of students.

Van Houten and Sullivan (1975) used an audio cueing system broadcast over a public address system to prompt teachers to praise students. Teacher praise was defined as “commendatory statements made by the teacher to a specified child contingent upon correct student academic behavior” (p. 198). Teacher behavior was observed but not
manipulated during baseline sessions; however the teachers were aware that the observers were present in their classrooms to assist with increasing praise rates. In the next phase, teachers were given counters and graph paper and were told to record and graph the number of praise statements they made during each observation session. They were told to maintain a praise rate of two per minute. The experimenter checked each teacher’s graph every day, but provided no feedback. In the third phase a beep was presented over the public address system in the classroom, and the teachers were instructed to look for an on-task student and to praise him or her whenever they heard a beep. These cues to praise were presented at a rate of two per minute. A return to baseline conditions followed this cueing phase. During the fifth phase the cues were reinstituted and were presented at a rate of three per minute. The final phase consisted of a return to baseline conditions. Their results demonstrated that teacher praise rates were not significantly altered by the recording phase, but that the cueing procedure was effective in increasing teacher praise. In addition, teacher praise rates did not decrease when the cues were removed. Two of the three participating teachers’ praise rates demonstrated increasing trends during their final baseline phases. The authors hypothesized that cueing was more effective than self-scoring because cueing involved a prompt to look for and attend to appropriate student behavior. The authors further noted that “it is possible that improvement in student performance [not systematically measured in this study] resulting from higher praise rates was sufficient to maintain teacher performance” (p. 200, words in brackets added). Although each teacher reported that student behavior improved with increased praise, it is also possible that as a result of being prompted to look for
appropriate behavior, teachers spent less time looking for and reacting to inappropriate behavior. Similar results were obtained in a replication by van der Mars (1987) with student teachers in physical education.

**Public Posting**

Public posting is a feedback-delivery strategy that involves displaying data on a target behavior in a location that is accessible to a study’s participants and to others in the participants’ environment. In the case of increasing teachers’ rates of positive statements, data on rates of positive statements are collected and transferred to a visual display.

Gross and Ekstrand (1983) demonstrated that publicly posting records of teachers’ praise rates resulted in increases in the number of praise statements made by teachers. In this study teacher praise was defined as “any positive comment directed by a teacher or aide to a student” (p. 128). The authors further specified that the teacher had to state more than just an exclamatory sound, and that positive comments directed to another staff member were not included. “This procedure was adopted to ensure that remarks recorded as instances of praise had the highest likelihood of also being identified as such by the children” (p. 129). Praise statements were measured with an interval recording system. Observers listened to audiotapes of instruction for twenty second intervals, and recorded the occurrence or nonoccurrence of teacher praise during the subsequent ten second intervals.

During baseline sessions the experimenter explained the importance of teacher praise and provided examples of contingent praise. No feedback was provided for recorded teacher praise during baseline sessions. During the public posting phase a 2ft. x 3ft. chart
was placed on the wall in each participating teacher’s classroom. This chart was a graphic representation of the intervals in which the participating teachers were observed praising the students; the chart included baseline rates. The chart was explained to each participant, and the experimenter updated the percent of intervals in which praise occurred each day. The third phase consisted of a return to baseline, and the public posting strategy was reinstated in the fourth phase. In the fifth phase, feedback to the participants was faded: the public posting chart was removed and teachers were told their percentage of intervals of praise on randomly selected days, approximately five times over a 3-week period. Following this condition, baseline conditions were reinstated. The study concluded with a follow-up phase 2 months later in which baseline conditions were in effect.

Results demonstrated not only that public posting was effective in increasing teacher praise rates, but also that it could be faded out successfully. One possible confound of this study is the training that occurred during the baseline conditions. Data demonstrated a variable but slightly increasing trend during the first baseline phase, indicating that explanations and examples of praise may have been sufficient for increasing praise rates for these participants.

This study presents a promising tactic for increasing teachers’ rates of praise statements, however the research on the effects of public posting on teachers’ praise is scant. Further studies are needed on this topic.

Instructions and Feedback

Another strategy that has been tested in several investigations is giving instructions, feedback, and praise to teachers, as teachers are told to do with their own students.
Cossairt, Hall, and Hopkins (1973) examined the effects of these elements for teachers. The researchers stated that since experimenters working with teachers often provide instructions on increasing praise rates in addition to using strategies such as cueing, the effects of these instructions themselves should be evaluated experimentally in the absence of other strategies. Two fourth-grade and one third-grade teacher participated in this study. Their years of teaching experience ranged from 2 to 4. Teacher praise was defined as “any positive or praise statement about student attending behavior” (p. 91). After each session in this study, the experimenter met with the teacher and provided feedback on the teacher’s praise rate for the most recent session. These sessions were audiotaped and the experimenter’s praise of the teachers was recorded.

The baseline conditions did not incorporate any instructions, feedback, or praise. Instructions were introduced in the next condition. The teacher was told to praise students who were attentive to instructions, and was given a written reminder of the effectiveness of praise at the start of this phase. In the next condition the teacher received feedback on her praise data for the most recent session. In the fourth phase social praise was added to the feedback, and after nine sessions, praise and feedback were delivered on an intermittent basis. One teacher participated in a return to baseline following this phase, and in a final “package” condition that included all three components. The results demonstrated that this strategy was most effective as a package. Instructions produced moderate, temporary increases in teacher praise. Feedback produced further increases, but feedback plus social praise was the most effective condition. A possible limitation of this study is that social praise was the only element of the package that was not assessed.
independent of the other two elements. The authors concluded that praise for teachers is essential for changing teacher behavior, and that since intermittent praise from the experimenter was effective in maintaining high rates of teacher praise, “the excuse that principals and supportive staff do not have time for the social reinforcement of teacher behavior is invalid” (p. 100).

In a similar study, Mace, Cancelli, and Manos (1983) also demonstrated that teachers delivered more contingent praise to their students when classroom consultants provided informational feedback and praise to the teachers for providing increased rates of praise. They defined contingent praise as “any positive verbal statement directed to the student regarding his performance during the instructional period” (p. 341) and also included “contingent reinforcing educational materials” as a dependent variable. Delivery of materials was defined as “teacher provides the student with access to academically related materials not otherwise available in the current instructional period on the basis of desired student performance” (p. 241).

Two conditions were employed across three participants: baseline and feedback plus praise. The feedback plus praise condition consisted of a conference between the consultant-experimenter and the teacher following each intervention session. In this conference the consultant showed the record of the data and provided praise for contingent praise and contingent materials. In sessions in which no praise or materials were delivered by the teacher, the consultant provided encouragement for the future use of contingent positive consequences. Results of this study demonstrated increased rates of
the delivery of praise and educational materials across all participants. Praise rates were higher than educational material rates.

Andrews and Kozma (1990) used a package of cueing, feedback, and praise to increase the praise that a fifth-grade teacher provided to her low-performing students. The researchers’ aim was to train the teacher to provide increased praise rates specifically to the lowest performing students in the class, as opposed to the class as a whole. The initial intervention condition included an audio cueing phase (cues were delivered via a micro cassette recorder). Cues were presented approximately every two minutes. In the next phase, the cues were included and the teacher was instructed to allocate two thirds of her praise to the lowest performing students in the class. In addition, the researcher delivered specific written feedback and praise to the teacher regarding her performance. This phase was followed by a removal of the instructions to praise the lowest performing students and a removal of researcher feedback and praise, and then by a condition that included praise for increased praise rates, regardless of the students to whom praise was addressed. The audio cues were present in the final two phases.

Results demonstrated that the teacher delivered no praise to the lowest performing students during the baseline condition and during the next to last condition, and that the condition that incorporated praise from the experimenter and instructions to praise the lowest performing group of students produced the highest rates of praise for these students. This study demonstrated that audio cueing can be used to increase specific kinds of teacher praise, and also that a package intervention for teachers, consisting of instructions, feedback, and praise, was most effective.
Instructions and feedback have been demonstrated effective in the modification of classroom teachers’ rates of positive interactions with their students. Such interventions can be time consuming and usually require the attention of a consultant or coach. Future research is needed on strategies to reduce the teacher’s reliance on another person.

Student Recruiting

Other researchers have trained students to recruit attention from teachers as a means of increasing teachers’ praise rates. Seymour and Stokes (1976) taught adolescent female students enrolled in a maximum-security institution to self-record work behavior. The girls were taught to monitor and record the following behaviors: hands working continuously; looking at work; doing the work set and not something else; not fiddling when getting something; not dawdling; and not talking out of place. The girls were also taught to cue staff to provide feedback on their work by pointing out their improvement. Results demonstrated that the participants did increase their work behavior and the number of cues, and also that once the cues increased, that staff praise to the participants also increased.

Stokes, Fowler, and Baer (1978) conducted a similar study in which they taught preschoolers to recruit positive feedback for their academic work. The children were first taught to identify “good work” and then to prompt or cue the trainer for positive evaluations when they considered their work to be good. Cues taught included, “Have I worked well?” and “How is this?” Results demonstrated that student work improved and that once students began recruiting feedback from their teachers, that teacher praise statements increased.
Alber, Heward, and Hippler (1999) also taught students with disabilities to recruit positive attention from teachers. Four sixth-grade students, three of whom who had been identified as having learning disabilities, participated. Two of the students received special education services in a resource room, and all of the students spent at least a portion of their day in general education classrooms. The purpose of the study was to assess the effects of teaching the students to recruit attention in a special education classroom; dependent variables included the number of recruiting responses the students emitted in general education classrooms, the number of praise statements the students received in their general education classrooms, the number of instructional feedback statements the students received in their general education classrooms, and the students’ productivity in their general education classrooms.

Student recruiting included three components and were only scored as correct when all three components occurred in sequence: each student was taught to raise his or her hand, wait for recognition from the teacher, and ask a question or make a comment to the teacher about his or her work. Student questions and comments included “How am I doing?” and “I don’t understand this one.”

Recruiting responses, teacher praise, and instructional feedback were measured as frequency counts; academic productivity was measured as percent work completed, and academic accuracy was measured as percent correct responses. Observers recorded the dependent variables in the general education classrooms only. Neither the general education teachers nor the participating students were told the purpose of the study.
The first author trained the resource room teacher to teach the students how to recruit teacher attention. Recruitment training consisted of explaining the rationale for recruiting teacher attention, and then role playing the steps of recruiting. Three students participated in two days of training, and one student participated in a third day.

After the training phase, the generalization programming phase (which included continuous reinforcement and intermittent reinforcement phases) began. The resource room teacher met with each student prior to homeroom and prompted each student to recruit positive teacher attention in his or her general education classroom. Students were prompted to recruit at least two times, but no more than three. Three students checked off a recording card each time they recruited teacher attention, and turned the cards in to the resource room teacher at the end of the day. The cards were printed with three boxes to check, and served to remind the students to recruit attention, but to recruit no more than three times.

Three students were primarily observed in their math class, and one was primarily observed in her social studies class. The three students who were observed in math class were also observed in a probe setting (social studies class) to check for generalization across settings. At the end of the day the resource room teacher checked in with the students; students were rewarded with one ticket (for an end-of-week classroom drawing) for each recorded recruiting response, and with a can of soda for recruiting teacher attention twice. If students did not recruit, they were encouraged to do so the next day.

To program for generalization, the end-of-day delayed reinforcement procedure was faded via an indiscriminable contingency: students were told that they would only check
in with the resource room teacher and receive tickets and soda on 2 days per week, which were randomly selected by the first author. During a subsequent maintenance phase, students did not check in with the resource room teacher.

Results demonstrated that three of the four students did recruit teacher attention at least once per 10 minutes for most sessions. Students also completed more independent work, and completed it with greater accuracy, in the generalization and maintenance conditions. All students who were observed in the probe setting (social studies class) increased their recruiting responses after training, and maintained higher rates during the maintenance phase than they had emitted during baseline.

In addition, teacher praise and academic feedback increased after students completed recruitment training. Teacher praise and academic feedback increased whenever student recruiting increased. Although teachers generally gave more instructional feedback than praise, praise rates did increase above baseline levels and corresponded with increases in student recruiting.

Craft, Alber, and Heward (1998) conducted a similar study with elementary students with developmental disabilities to recruit teacher attention in the general education classroom. Recruitment training consisted of instruction and role playing (which lasted for two days), followed by daily (morning) prompts to recruit and end-of-the-day checks and rewards. Baseline data demonstrated that only one of four participating students received any praise from the general education teacher. After training, all students demonstrated increases in recruiting and teachers demonstrated increases in praise. In
addition, students’ completion and accuracy scores on spelling worksheets increased once the training phase was implemented.

Wallace, Cox, and Skinner (2003) taught a 10-year-old male student with mild mental retardation to recruit teacher attention as part of a two-component intervention to increase independent academic work productivity. This student had difficulty completing math worksheets accurately. The sheets generally contained 30 subtraction problems, and the student usually completed about five problems accurately.

The intervention consisted of giving the student five or six problems at a time, and telling him to raise his hand or bring the sheet to the teacher when he was finished. When the student turned in the completed worksheets, the teacher praised his work completion and gave him another worksheet with five or six more problems. This procedure was followed until six sheets had been completed, or for 20 minutes.

Intervention data indicate increases in the student’s work completion and accuracy, replicated across an ABAB design. Teacher disapprovals decreased during the intervention phases, and approvals increased to almost 10 times the number observed in baseline.

Recruitment training has been shown to produce increases in recruiting responses, teacher feedback, and work productivity. These studies have also demonstrated that this is an effective strategy that not only increases positive student-teacher interactions, but also provides students with disabilities with an important skill that they can use in regular and special education settings, and in vocational settings.
Goal Setting

Goal setting has been used to change a variety of behaviors in a range of settings (Langeland, Johnson, & Mawhinney, 1998; Ward & Carnes, 2002). Goal setting involves setting a target number for a particular behavior, and then comparing actual performance with the target number. Goal setting can be used to increase or decrease behavior.

Two studies on the use of goal setting to change teacher behavior were conducted by Martens, Hiralall, and Bradley (1997) and Sharpe, So, Mavi, and Brown (2002). Martens et al. used goal setting to increase a kindergarten teacher’s praise statements to students with disabilities, and Sharpe et al. used goal setting to increase effective teaching practices by four undergraduate students in a physical education specialist certificate program.

Martens et al. (1997) measured teacher praise using event recording and measured two students’ schoolwork (attending to task), attending to instruction, and responding aloud behaviors with a 15 second momentary time sampling procedure. A multiple baseline across students design was used to evaluate the effects of the teacher’s setting a goal for the number of praise statements she was to provide to each student during each 30-minute observation period.

During baseline, the teacher conducted her class as she normally did. No feedback was provided to the teacher during baseline. Praise rates were low for both students, and decreased over the course of the baseline phase. During the goal setting phase the teacher identified four behaviors to target for increase for both of the students, and she set a goal for increasing the number of praise statements she delivered to each student. At the start
of each session in the goal-setting phase one of the observers gave the teacher a note indicating the number of praise statements she had delivered during the previous observation. The feedback note also included praise for the teacher if she had met her goal. The goal was constant throughout this phase.

Baseline data indicated highly variable levels of appropriate behaviors by both students and low, but decreasing numbers of praise statements by the teacher. Upon starting the goal setting condition, levels of student appropriate behaviors became high and steady, and teacher praise statements increased above baseline levels.

Sharpe et al. (2002) used a goal-setting procedure to increase the instructional behaviors of four undergraduate students in a physical education teacher training program. Undergraduates were observed in their placement settings. The two settings were middle school physical education classes and university laboratory classes with undergraduate peers who took the role of middle school students. Students attended placements for a 12-week semester. All four participants were observed at least 12 times per semester, for two semesters; two participants were observed in both settings, and the other two participants were each observed in one setting. Participants chose which placement to complete first based on their own scheduling needs.

Instructional behaviors targeted for increase included interpersonal and organizational behaviors and instruct/observe/feedback sequences with students. An ABACA design was used with two participants to evaluate the effectiveness of feedback and goal setting; A represented baseline conditions, and B and C represented the feedback and goal setting conditions in either the middle school classes or the laboratory
school classes for each participant. The other two participants were only exposed to one teaching setting, but afterwards participated in an extended maintenance condition during which all feedback and goal setting was discontinued and baseline conditions resumed.

The intervention consisted of 15-minute individual meetings with a graduate student observer to review performance, and occurred immediately after each observation. During these talks, the participants reviewed what they thought had gone well during their lessons and what had been the most challenging aspects of their lessons. Observers then presented the data from the lesson, and the observer and the teacher set goals for up to three behaviors to improve at the next observation.

Data demonstrated increases in instructional sequences and decreases in organizational and interpersonal behaviors for participants in the middle school-based goal setting condition. The university laboratory school-based goal setting condition produced small, short-lived improvements in participants’ behavior.

Although this study does not report specifically on increasing teacher praise, the increases in complete instructional sequences demonstrates that goal setting is a promising procedure for increasing instructional behavior, including praise statements.

**Self-Scoring and Self-Evaluation**

Self-scoring and self-evaluation have been advocated as promising methods that teachers can easily use to improve their skills in the classroom (Gunter & Reed, 1996). Such procedures have sometimes also been called self-recording, self-assessment, self-monitoring, and self-evaluation. For the purposes of this review, self-scoring refers to
quantification of one’s behavior (e.g., a tally), and self-evaluation refers to other forms of data collection, such as making a checkmark or recording “yes” or “no” on a checklist.

Self-scoring and self-evaluation have been implemented by several researchers to change a variety of teaching behaviors. Hoover and Carroll (1987) used a self-scoring and self-evaluation checklist to improve the reading instruction of 53 elementary school teachers (2 kindergarten, 16 first grade, 8 second grade, 9 third grade, 10 fourth grade, 5 fifth grade, 2 sixth grade, and 1 seventh grade).

As part of an ongoing training program in reading instruction, teachers were asked to make pre-intervention audiotapes of two different reading groups. At the next training meeting (two weeks after the first training), the university consultant delivering the training instructed the teachers in the use of a checklist that included 10 aspects of teacher behavior. Five items were scored “yes” or “no”, and five (including “specific academic praise”) were scored with frequency counts. Desired frequency count behaviors were listed next to their undesirable counterpart behaviors; “specific academic praise” was listed next to “general praise.” Teachers practiced using the checklist to score the behavior of other teachers from video and audiotapes, and were instructed to score their own tapes using the checklist and to then identify behaviors that they wanted to increase or decrease. They were told to practice these changes before making and scoring two post-intervention tapes.

Data collection was conducted by scoring 80 randomly selected audiotapes (40 pre-intervention and 40 post-intervention). Analysis of the data across the scored tapes revealed that four of the five yes/no and all five of the frequency count teaching
behaviors were higher after the teachers had employed the self-scoring checklist. The largest change across all recorded behaviors was the increase in specific academic praise, which occurred more than 3 times more often on post-intervention tapes than on pre-intervention tapes.

Richman, Riordan, Reiss, Pyles, and Bailey (1988) implemented a self-evaluation procedure to increase the on-task behavior of staff in a residential setting. Staff were given activity cards to carry with them during their shifts to help them to determine activities to implement with clients during scheduled time periods. Staff members were instructed to initial their cards upon completion of an activity. Results demonstrated that this procedure was effective in increasing on-task and on-schedule behavior, and that the addition of supervisor feedback produced further improvements.

Belfiore and Browder (1992) used a self-evaluation procedure to increase the accuracy of teachers’ data-based decision making. The experimenters gave the participants a data sheet on which to record salient information regarding instructional decisions. This information included the decision date, the decision rule, the specific change that was made, the current mean, the previous mean, and the success of previous decision. After implementation of this form, teachers demonstrated high, stable percentages of accurate decision-making.

Several studies have addressed self-scoring as a tool for increasing teachers’ praise rates in their classrooms. Cusimano (1987) used a pretest-posttest control group design to teach eight elementary physical education teachers to self-score verbal teaching behaviors. The dependent variables were positive specific feedback, corrective specific feedback,
and acceptance of students’ skill performance ideas. Pretest and posttest recording took place during each participant’s skill instruction during physical education lessons; each participant recorded three 15-20 minute lessons for the pretest and three 15-20 minute lessons for the posttest. Training in self-evaluation consisted of two inservice sessions, which also included discussion on goal-setting strategies. For the 5 weeks following the inservice training the participants audiotaped and tallied the occurrence of each dependent variable. The experimenter provided written feedback and a graph of performance on a weekly basis. Results demonstrated a significant increase in positive specific feedback and in corrective specific feedback for the experimental group.

Sutherland and Wehby (2001) also used a pretest-posttest control group design in their evaluation of the effects of self-scoring with teachers of students with emotional and behavioral disorders. Ten teachers participated in each group. All participants volunteered to take part in the study, and were compensated for their participation ($100 for teachers in the control group, and $150 for teachers in the treatment group).

All participants taught students in grades K-8. There were 216 students in the 20 classrooms of the participating teachers. Students had previously been identified as having emotional disturbance, a learning disability, or mental retardation, or were otherwise categorized. Student ages ranged from 5 to 15 years.

Pretreatment sessions consisted of a minimum of 90 minutes of observation during which an observer recorded the occurrence of nine teacher behaviors (behavior-specific academic praise; behavior-specific social praise; non-behavior specific praise; reprimands for academic behavior; reprimands for social behavior; group opportunities to respond;
individual student opportunities to respond; academic talk; other talk) and student correct responses. Observation sessions were conducted during the first 15 minutes of teacher-led instruction across subject areas. One session was conducted per day.

After the pretreatment phase all teachers in the treatment group met individually with the experimenter to discuss examples and definitions of effective praise. In this meeting each participant was also trained to use a cassette recorder and count and record praise statements while listening to a 5-minute audiotaped sample of him or herself engaged in classroom instruction. Finally, participants in the treatment group were instructed to set a goal praise rate, to graph their praise rates daily, and to use self-praise for attempting to increase praise. After this meeting, a 2-week “treatment” phase began, during which teachers turned in their audiotapes to a research assistant daily and maintained graphs of their praise rates. Direct observational data were collected for at least six 15-minute sessions in this phase.

Teachers in the control group also met individually with the experimenter, who gave them a cassette recorder and taught them to use it to record themselves during classroom instruction. These teachers also exchanged tapes daily with a research assistant, and data were collected on these teachers’ behaviors for at least six 15-minute sessions during the treatment phase.

Approximately one month after the conclusion of the treatment phase, four 15-minute observation sessions were conducted with each teacher to assess the maintenance of any changes produced during the treatment phase.
Results were reported for total teacher praise (three categories collapsed into one), total reprimands (categories were collapsed), total opportunities to respond (group and individual were combined), student correct responses, academic talk, and other talk. Data demonstrated that the treatment group’s mean number of total praise statements per minute was approximately 2.5 times higher than the mean pretreatment rate, however this effect was not maintained a month after treatment was terminated, although the maintenance rate of total positive statements was still higher than the pretreatment rate. Participants in the control group demonstrated a decrease in total positive statements during treatment, and a small increase over treatment rates during maintenance observations.

Another noteworthy result of Sutherland and Wehby’s (2001) study was a decrease in the rate of total reprimands by the teachers in the treatment group. Reprimands occurred at a mean of .43 per minute during pretreatment, .3 during treatment, and .19 during maintenance, despite the fact that reprimands were not addressed by the self-scoring protocol used with the treatment group.

Sutherland and Wehby (2001) speculate that teachers in the treatment group became more aware of their teaching behavior as a result of the self-scoring procedure, and as a result attempted to improve their instructional practices. Increased awareness of teaching behavior could be responsible for an increase in praise and a decrease in reprimands, however since reprimands were not self-recorded by the teachers, it remains to be seen what component of the self-scoring intervention was responsible for the changes observed.
Sutherland and Wehby (2001) wrote that a possible explanation of the failure of the self-scoring intervention to effect lasting changes in treatment group teachers’ behavior is that the “classroom effects that occurred (or did not occur) as a result of increased praise may not have been sufficiently reinforcing for the teachers to maintain their increased rates” (p. 169). Any observed increases in appropriate student behavior would probably not have functioned as a reinforcer for teacher praise, however, because such changes are generally too far removed in time from the occurrence of the teacher’s praise behavior.

Although Sutherland and Wehby (2001) demonstrated that a self-scoring procedure could be effective, the most significant limitation of their investigation was the limited amount of time participating teachers spent with the experimenter. A potentially effective variation on their procedure would incorporate more contact with the experimenter; such a change would perhaps have produced outcomes with greater durability. In addition, increased time with the experimenter could have led to the development of individualized strategies that the teachers could use to increase their praise rates.

Another useful variation would be to assess the effects of self-scoring with a single subject, instead of a groups-comparison design. A single-subject design with a smaller number of participants would allow for more observation sessions, closer contact with the data on a daily basis, and the opportunity to make changes in the self-scoring procedures if they are not effective in changing teachers’ behavior. The analysis of behavior change over time within and across conditions would contribute to the analysis of the effectiveness of a self-scoring procedure such as the one described by Sutherland and Wehby (2001).
Individually Determined Interventions

Several studies have demonstrated the effectiveness of a “participative management” (Burgio, Whitman, & Reid, 1983) approach to changing teacher and direct-care staff behavior. Clark and Macrae (1976) wrote that teachers in their study who were allowed to self-select a training package demonstrated greater gains in classroom teaching skills than did teachers who used an “imposed” package. Self-selected interventions tend to have higher face validity and are more likely to be rated as acceptable by the participants who select and implement them. The use of the “self-selection” procedure might be a powerful addition to a self-scoring intervention.

Summary of Literature

Research on the effects of teacher praise and on strategies to increase teacher praise demonstrates several important points:

(1) Teacher attention often functions to reinforce student behavior, regardless of whether that attention is positive or negative. When teacher praise and attention are used to increase appropriate student behavior, on-task behavior and learning increase.

(2) The reactive effects of self-scoring may be beneficial components of treatments to improve teachers’ skills. Self-scoring is also a procedure that can be implemented without extended and intense supervision, which makes it an economical choice for intervention in terms of time and personnel.

(3) Investigations on the use of self-scoring to increase teachers’ praise statements are promising, however further study is needed in order to develop a sound literature base
on which teachers and teacher trainers could begin to base their practices in applied settings.

The present study combined research on self-scoring with previous research on a variety of strategies to increase teacher praise. Given that research has demonstrated that self-selected interventions may be more effective than imposed interventions, this study employed both self-scoring and, as needed, self-selected, individualized interventions to increase teacher praise during classroom instruction.

Research Questions

1. What are the effects of self-scoring the number of positive statements from audio and videotapes of instruction on teachers’ rates of positive statements delivered during subsequent instruction?

2. What are the effects of an intervention package consisting of self-scoring the number of positive statements from audio and videotapes of instruction and individually determined interventions on teachers’ rates of positive statements delivered during subsequent instruction?

3. What are the effects of self-scoring the number of positive statements from audio and videotapes of instruction on teachers’ rates of specific and varied positive statements, and negative statements, delivered during subsequent instruction?

4. What are the effects of an intervention package consisting of self-scoring the number of positive statements from audio and videotapes of instruction and individually determined interventions on teachers’ rates of specific and varied
positive statements, and negative statements, delivered during subsequent instruction?

5. What are teachers’ perceptions about the intrusiveness and effectiveness of self-scoring the number of positive statements from audio and videotapes of instruction and/or an intervention package consisting of self-scoring the number of positive statements from audio and videotapes of instruction and individually determined interventions?

6. What are current and future teachers’ perceptions of “before intervention” and “after intervention” videotaped samples of participating teachers’ instruction?
CHAPTER 3

METHOD

The experimental procedures, data collection methods, and data analysis as they are related to the research questions are described in this chapter. Description of the participants, setting, procedures, measurement, experimental design, dependent variables, and social validity are detailed.

Participants

Four teachers employed at a large Midwestern charter school participated in this study. This elementary school was in its fifth year of operation. The school employed Direct Instruction curricula for math (Connecting Math Concepts: Engelmann, Carnine, Engelmann, & Kelly, 1994), reading (Reading Mastery: Engelmann & Bruner, 1995a, 1995b), and language instruction (Language for Learning: Engelmann & Osborne, 1998). The school served more than 700 students in grades K-5. The school was in Academic Watch status during the course of this study.

Three teachers participated in the primary component of this experiment, and a fourth teacher participated in a secondary component. These two components are explained in greater detail in the experimental design and procedures sections.
Prior to soliciting participation in the study, the experimenter, who was employed as a graduate assistant at the school, conducted observations in all kindergarten, first, and second grade classrooms to assess teachers’ use of positive statements. Potential participants were selected on the basis of several criteria. Candidates for participation were selected to receive informational letters and consent forms for the study if they demonstrated a low (for the purposes of this study, defined as less than .5 per minute) rate of positive statements during an observation of classroom instruction, or if they had specifically requested assistance with improving their classroom management skills. Eight informational letters and consent forms were distributed to candidates, and five were returned. One interested participant was eventually excluded when her rate of positive statements increased during baseline data collection. All participants had demonstrated a consistent record of attendance. One participant had earned a M.Ed. in English education; however graduate degrees and teaching certification were not required for participation.

The general procedures of the investigation were explained to the participants, and were provided in writing via a consent form (see Appendix A). Approval to conduct the study was obtained from the board of directors of the school and the Institutional Review Board of The Ohio State University before the study began.

Teacher 1

Teacher 1 was a first grade teacher. She started work as an instructional aide at the school 5 months prior to this study, and had been promoted to a teaching position after 2 months of employment. Prior to working at the school she had worked in a daycare
setting for 1.5 years. Upon her promotion to a teaching position she told the experimenter that she was interested in assistance with classroom management. She had earned a B.A. in cross-disciplinary studies.

The 27 students in Teacher 1’s class were considered to be at an average level compared to other first grade classes at the school. The audio recording of Teacher 1’s instruction for the study took place daily, immediately after morning announcements and the recitation of the pledge of allegiance, during a teacher-led review of previously learned material and a preview of the new material for the day. Recording commenced at approximately 9:15 a.m. and lasted until 9:40 a.m. The videotaping for generalization sessions took place once per week during math lessons, which began at approximately 2:00 p.m.

Teacher 2

Teacher 2 was a second grade teacher, and the only participant who possessed a graduate degree (a M.Ed. in English education) and teaching certification. She had taught at the school for three years, and had previously taught at another school for one year. She had requested information from the experimenter on improving her classroom management skills. Her class of 24 students was considered to be at a high level compared to other second grade classes at the school.

The audio recording of Teacher 2’s instruction for the study took place daily from 10:45 a.m. until approximately 11:10 a.m., during an independent work period when students worked on catching up on a daily review material in reading, language, science, and math. The videotaping for generalization sessions took place once per week after
morning announcements and the pledge of allegiance and during her daily review and preview of instruction, from approximately 9:15 a.m. until 9:40 a.m.

Teacher 3

Teacher 3 had been teaching for 11 years; she had taught at the school for three years. She possessed a B.S. in marketing and management. Her class of 20 students was considered to be at a low level compared to other third grade classes at the school. Upon referring one of her students to the school’s Intervention Assistance Team (IAT), she had indicated to the experimenter that she was interested in assistance with classroom management.

The audio recording of Teacher 3’s instruction took place at the end of the school day at approximately 2:30 p.m., during math, social studies, and/or spelling instruction. The videotaping for generalization sessions took place once per week during her daily morning review and preview of material in math, language, social studies, science, and reading, for about 25 minutes starting at approximately 9:15 a.m.

Teacher 4

Teacher 4 was the school’s music teacher. He had been teaching for 5 years, 2 at the school. He had earned a B.A. in music education, and was pursuing teaching certification. He had previously requested classroom management assistance with his second grade students. The audio recording of Teacher 4’s instruction took place during a second grade class that met on Mondays, Tuesdays, Thursdays, and Fridays; recording began at approximately 1:50 p.m. and ended at about 2:15 p.m. Because students only attended music once per week, different students attended this class each day.
Setting

The study took place at the school. Audio and video recordings, meetings, and self-scoring took place in classrooms, which consisted of student desks, two teacher desks, a computer, and shelves containing instructional materials. Recordings occurred at various times during the school day, including individual seat work and whole-class activities, but audiotaping always occurred at the same time each day for individual participants. The experimenter scheduled individual recording sessions and meetings in collaboration with each participant. Videotaping was conducted simultaneously with audiotaping once per week; a digital video camera mounted on a tripod was used.

In addition to the daily recording sessions, generalization probes were conducted once per week during all phases for Teachers 1, 2, and 3, in order to assess generalization of any changes in teachers’ rates of positive and negative statements across settings, subjects, and times of day. Generalization probes were conducted at the same time of day, but not on the same day, once per week. Participants were aware that this recording would take place, but did not know the specific times scheduled for observation. Generalization recording was accomplished via small wall-mounted wireless video cameras that transmitted to receivers connected to videocassette recorders in classrooms adjacent to the participants’ classrooms. Generalization probes were different from regular sessions in that participants were not able to detect when recording was taking place and also in that wireless microphones were not used.
Experimenter

The experimenter was enrolled in the Ph.D. program in Special Education at The Ohio State University. In 1995 she received a B.A. in Psychology from Smith College in Northampton, Massachusetts. In 2000 she received a M.A. in Behavioral Disorders from Columbia University Teachers College in New York. She was a certified Special Education Teacher in the state of New York and a nationally Board Certified Behavior Analyst. For the duration of the academic year during which this study was implemented she was employed as a 50% FTE graduate assistant at the elementary charter school where the study took place. Her responsibilities included assisting in the coordination of services for children with autism who attend the school and providing behavior management and intervention assistance to teachers.

As a result of her assistantship responsibilities, the experimenter spent time conducting formal and informal observations in each kindergarten, first, and second grade classroom in the school in the months prior to this study. She was also a member of the school’s Intervention Assistance Team. She was therefore familiar with each of the teachers who subsequently participated in this study. Prior to the study all four participants had requested assistance in their classrooms, and the experimenter had conducted several observations of teacher and student behavior in each class. In addition the experimenter had made recommendations to Teachers 3 and 4 to increase the positive attention they provided to their students, however their rates were sufficiently low at the start of this investigation to warrant their participation.
Data Collectors

The experimenter served as the primary data collector. Two undergraduate students and one graduate student in special education served as secondary data collectors. They were recruited via the program’s e-mail list and via announcements in class. They were compensated with independent study credit hours for their assistance with this investigation.

Materials

Materials included electronic and manual recording equipment. Electronic equipment consisted of a digital video camera, audio and video cassettes, and three sets of wireless audiotaping equipment, each of which included a microphone and receiver (Radio Shack FM Wireless Microphone System, Cat. No. 32-1221B) and an audio cassette recorder. Electronic equipment also included three sets of wireless cameras and receivers (X10.Com XCam2 Wireless Color Camera, Model XX16A; X10.Com Video Receiver, Model VR31A-C), and combination television/video cassette recorders. Manual recording equipment included data sheets (see Appendix B for blank and sample completed data sheets), an audio cassette player, writing implements, and a digital timer.

Definition and Measurement of Dependent Variables

Definitions

All dependent variables were measured with a frequency recording system and were converted to rate (number per minute), with the exception of repeated statements. Repeated statements were counted and summarized as a percentage of each category of
statements. In addition, the respective time index for each statement was recorded. See Appendix B for an example of a completed data sheet.

Generic positive statements. Generic positive statements consisted of audible statements made by the teacher that were addressed to students and referred to one or more than one student’s behavior or work products as desirable or commendable. Examples included, “I’m proud of you!” and “Great job, everyone.” Generic positive statements did not refer to specific behavior and may or may not have indicated student names. Neutral statements only indicating the correctness of an academic response (e.g., “Okay”, “correct”) were not counted. Statements made to other adults in the room were counted if they were loud enough to be heard by the students and if these statements made direct reference to student behavior or work products (e.g., “Aren’t you impressed at how quietly my students are working today?”). Statements that would not be recorded included incomprehensible or inaudible statements or those that were not related to student behavior (e.g., saying “Thanks for dropping off my attendance forms” to a co-worker). A series of positive comments that did not specify behaviors was counted as one statement unless there was a separation of at least 2 seconds between comments (e.g., “good, good, good, I’m so impressed,” stated when reviewing three or four students’ work, or one student’s work, would be counted as one statement).

Behavior-specific positive statements. Behavior-specific positive statements were similar to generic positive statements, with the exception that behavior-specific positive statements made explicit reference to an observable behavior (e.g., “Thank you for putting your pencil away”). Specific positive statements could refer to general classroom
behavior (e.g., “You did a great job walking back to your seat quietly”) or academic behavior (e.g., “That was a super smart answer!”). Specific positive statements were separated by 2 seconds or by differentiation of the behavior that was targeted for praise. In other words, if a teacher described or named a desirable behavior and then listed multiple students who were demonstrating the behavior, this would be recorded as one statement. For example, “Marissa, Tony, and Mark, you did a great job of returning your materials when you were finished with them.” would be recorded as one specific positive statement. If the teacher commented positively on several behaviors, each comment about a different specific behavior was counted as one specific positive statement. For example, “Jade, you did a great job cleaning up so quickly, Charles, thanks for putting the workbooks away, and class, I appreciate that you lined up quietly.” would be recorded as three separate and specific positive statements regardless of the interval between the end of one comment and the start of the next.

Negative statements. Negative statements expressed disapproval of a student’s behavior (e.g., “I don’t like that behavior in my classroom”) and/or included a prohibitive statement (e.g., “Stop” or another instruction to cease engaging in a behavior). Rhetorical questions were also included (e.g., “Why are you out of your seat?” “Why are you yelling?”). Negative statements could refer to managerial or academic behavior. Neutral statements indicating only that an academic response was incorrect were not counted, for example, “No, try again.” Negative statements were separated from one another by at least 2 seconds or by specification of different behaviors. Statements made to other adults in the room were counted if they were audible enough to be heard by the students and if
these statements made direct reference to student behavior or work products (e.g., “I don’t know what’s wrong with these kids today; they just can’t get it together”). Statements to other adults that would not be recorded included inaudible or incomprehensible statements, or negative statements directed to an adult that are not related to student behavior (e.g., “Can you believe we have another staff meeting after work today?”).

Repeated statements. The variety of statements was assessed by calculating the percentage of statements within each category that were repeat statements during each session. Each statement made by each participant was recorded verbatim the first time it occurred during a session, along with the time index at which the statement occurred. Each subsequent use of that statement was coded as a “repeat” by recording only the time index at which the statement was observed (See Appendix B for an example of a completed primary and secondary observer data sheet). For example, if a teacher responded to a correct student answer with the statement “Great!” the first instance of “Great!” was recorded as a generic praise statement. Every use of “Great!” thereafter was tallied in a column next to the original statement by recording only the time index at which the repeat was observed. If a teacher used the word “Great” as part of varied specific praise statements (e.g., “Great job finishing your work quickly,” “Great work on your journal assignment”) these were counted as two different (i.e., not repeated) specific positive statements. At the end of each session the total number of generic positive, specific positive, total positive, and total negative statements was calculated, and the total
number of repeated statements was divided by the total number of statements to produce a percentage of repeated statements within each dependent variable.

Measurement and Recording

Audio recording took place daily. Audiotaping lasted for a minimum of 15 minutes, and a maximum of 25 minutes. With the exception of the final phase for Teacher 1, Teacher 2, and Teacher 3, all sessions were audiotaped, and one audiotaped session per week was simultaneously videotaped. The video camera was placed so as to record activities that occurred throughout the room. In addition, one 25-minute generalization video recording was conducted each week via the wireless camera mounted in each of the three primary participants’ classrooms. The videotape recorder connected to each camera’s receiver was programmed to record once per week at a time unknown to the participants. For audiotape recording the teacher wore a small wireless microphone that transmitted to a receiver that was connected to a cassette recorder.

The first 15 uninterrupted minutes of each audiotaped regular session or videotaped generalization session were used for data collection. Data collectors used cassette players and digital timers to listen to and time the taped sessions. If any interruptions occurred, data collection was suspended (i.e., the timer was paused) until regular classroom routines recommenced. Interruptions included a phone or intercom call to the classroom that required the teacher’s attention, a classroom visit from a parent, administrator, or school employee that required the teacher’s attention, an announcement that caused the teacher to stop instruction, or a fire drill or other drill or actual emergency that required the class to leave the room. Recording of the dependent variables was accomplished by
simultaneously starting the audiotape (or videotape) of each session and a stopwatch. Both the audiocassette player and the stopwatch were stopped when positive or negative statements were heard, and the statements and the time index shown on the stopwatch were recorded. For subsequent uses of the same (i.e., repeated) statements, only the time index was recorded. At the end of 15 minutes, data collectors counted the number of generic positive, specific positive, total positive, and total negative statements for each session, along with the respective rates of each kind of statement and the percent repeats of each kind of statement. All data were recorded on the experimenter data sheet found in Appendix B, and were subsequently copied to a Microsoft® Excel spreadsheet for summarization and graphing.

Procedures to Enhance Believability of Data

Interobserver Agreement

Secondary data collectors were trained to observe and record all categories of positive and negative statements made by participants. Secondary data collectors used the same data sheets that the experimenter used, and recorded all positive and negative statements made along with their respective time indexes.

Three observers were trained prior to collecting data independently. The first training session involved the secondary data collector and the experimenter listening to an audiotape of a participant. The experimenter demonstrated the use of the cassette player, the stopwatch, and the data sheet. The experimenter gave the secondary data collector a set of data collection materials (cassette player, data sheets, stopwatch, written instructions for scoring, and audiotapes) and instructed her to score four sessions on her
own after the training session. Each secondary data collector returned the audiotapes and the completed data sheets to the experimenter, who then compared the secondary data collector’s data sheets with her own. Because the secondary observers transcribed the target statements, the time indexes were not used in determining agreement with the primary data collector. Agreement of frequency was counted on an occurrence by occurrence basis according to the transcript of the statements. Discrepancies in the recording were discussed with the data collectors to clarify the observed behavior and the behavioral definitions. Training of secondary observers continued until agreement scores were above 85% across target behaviors for three consecutive sessions. If agreement with any one data collector was below 85% for two consecutive sessions, the experimenter and the data collector reviewed and discussed examples and non-examples of the definitions of the dependent variables.

Interobserver agreement (IOA) of frequency was calculated by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying by 100 to produce a percent agreement score. IOA was assessed on 40.2% of all sessions, across phases.

*Treatment Integrity*

Fidelity of the application of the independent variable was assessed by visual examination of the participants’ self-scoring graphs. First, the dates of the recorded self-scoring sessions were compared to the dates of the onset of the intervention conditions to determine if the target number of self-scoring sessions were conducted each week (e.g., four sessions during the first week, reduced by one per week until meetings were
occurring once per week). The number of actual meetings was divided by the target number of meetings and the resulting number was multiplied by 100 to produce a percent of target meetings score. Next, each session recorded on each participant’s graph was checked for three elements: the recording of positive statements and their respective time indexes on each self-scoring data sheet, the correspondence of the number or rate of positive statements recorded on the data sheet with the number recorded on the graph, and the goal for the next self-scoring session recorded on the graph. Therefore, three possible points could be counted for each session held. The number of actual points observed was divided by the total possible, and the resulting number was multiplied by 100 to produce a percent of meeting elements score.

Checks were also conducted to ensure that data were collected from wall mounted cameras when specified by the conditions in place, and that any supplemental strategies were used as prescribed.

Social Validity

To assess the social validity of the goals, procedures, and outcomes of this study, several components were incorporated to assess the opinions of the participants and of current and future teachers. A modified version of the Treatment Acceptability Form - Revised (TARF-R) (Reimers & Wacker, 1988) was used to assess participants’ opinions of the effectiveness and acceptability of the study (see Appendix G). In addition, sample segments of videotapes that included both baseline and intervention clips were shown to current and future teachers for their assessment of the validity of the results (see Appendix H). This group comprised undergraduate and graduate students enrolled in
special education programs at a local university and at a college in the northeast. The students were recruited by their professors, who were informed of the study by the experimenter. None of the individuals who completed the survey were told the purpose of the study before they filled out the survey.

Experimental Design

A multiple baseline across participants design was used to assess functional relationships between self-scoring and teachers’ rates of positive and negative statements. Data collection began on all participants at about the same time, and was initiated as soon as possible after each participant had returned a signed consent form. Each phase of the intervention was implemented in a staggered fashion such that changes were made for only one participant at a time while conditions for the other participants were held constant. No changes were made until data demonstrated stability. The experimental logic of this design is that changes in the dependent variable occur when – and only when – changes are made in the independent variables manipulated by the experimenter.

The procedures detailed in the sections that follow were implemented with Teacher 1, Teacher 2, and Teacher 3. Procedures for Teacher 4 occurred in the same order, however the baseline phase was preceded by an extended phase consisting of weekly videotaping. This preliminary condition was conducted while the primary participants (Teachers 1, 2, and 3) participated in the baseline and intervention conditions. Once the primary participants had all begun self-scoring, Teacher 4 began daily audiotaping. When his data demonstrated stability, self-scoring was implemented. Due to time constraints, training (see section 3.7.2) was not conducted with Teacher 4.
Procedures

Baseline

In this condition participants were audiotaped daily and videotaped weekly. They received no feedback on their behavior while teaching and did not have access to the tapes for scoring or review. Prior to the start of each session the experimenter prepared the recording system (audio and/or video). The experimenter remained in the room for all sessions in which video and audio recording occurred simultaneously, and for audiotaped sessions as her schedule permitted. After the session ended, the experimenter disassembled the recording devices and removed them from the classroom. Data were collected on all categories of positive and negative statements. No data were collected in the classroom while video or audio taping was occurring.

Training

Prior to implementing the self-scoring condition the teacher was trained to score another person’s positive statements from a recording. The experimenter served as the trainer. A sample videotape of a teacher at the charter school who was not participating in the study was used for training participating teachers to score positive statements (generic and specific were grouped together and negative statements were not scored). Training did not begin until baseline data on positive statements demonstrated stability or a counter-therapeutic (i.e., descending) trend. Intervention did not begin until training had been completed successfully and data on positive statements again showed stability or a counter-therapeutic trend. Training consisted of scoring all positive statements (i.e., generic and specific were grouped into one category) only; negative statements were not
scored. The experimenter continued to collect data on separate categories of positive and negative statements (as in the baseline condition). Training was conducted with Teacher 1, Teacher 2, and Teacher 3 only.

The teacher and the experimenter practiced scoring positive statements by observing a 5-minute segment of the training video and transcribing the positive statements that the teacher made and the time at which each statement was made, according to the counter on the VCR (see the Participant Data Sheet in Appendix B). Observation data for training and self-scoring phases were compared for agreement. Agreement of frequency was by calculated on an occurrence-by-occurrence basis. The participating teacher’s occurrence measurement of positive statements was divided by the occurrence measurement of the experimenter. The result was multiplied by 100 to produce a percent agreement score. Agreement was always 100%.

**Self-Scoring**

During the self-scoring phase, participants met with the experimenter to review a 5-minute segment of the audiotape. Meetings were initially scheduled to take place on five days during the first week, but at participants’ requests the number of meetings during the first week was reduced to four. Meetings occurred on three days during the next week, and two days during the third week, until the experimenter and the participant were meeting once per week. The experimenter maintained a 3-ring binder that held all data sheets and graphs for all participants. The participating teachers filled out their self-scoring sheets and graphs, but the experimenter was in possession of all self-scoring materials used by the teachers.
To determine the 5-minute sample that would be self-scored, the participant advanced the audiotape and stopped wherever he or she desired. Participating teachers only scored positive statements; no differentiation was made between generic positive and specific positive statements, and negative statements were not recorded. Otherwise, participants scored the 5-minute samples in the same way the experimenter and secondary data collectors scored the 15-minute sessions. Participants listened to the audiotape and transcribed the positive statements that they heard; they also recorded the time index on the stopwatch (all four participants requested that the experimenter hold the stopwatch during the self-scoring sessions). In order to avoid influencing the participants’ self-scoring, the experimenter did not stop the stopwatch after a positive statement unless the participant had stopped the cassette player. If a participant had missed a positive statement, the experimenter would have stopped both the stopwatch and the cassette player, rewound the cassette, and asked the participant to listen again. This positive statement would not have been recorded on the participant’s data sheet. If a participant asked whether or not a statement met the definition of a positive statement, the experimenter stated the definition and told the participant that it was his or her decision to record the statement. Participants did not miss any of their positive statements and were more likely to ask for clarification on statements that did not meet the definition than on statements that did. None of the participants demonstrated any difficulty in scoring their tapes, including Teacher 4, who did not participate in the training session with the videotape of the non-participating teacher. See Appendix B for a sample of a completed participant data sheet.
Scoring continued until five minutes had elapsed on the stopwatch. After scoring the audiotape, the experimenter removed the audiotape of the scored session and replaced it with a blank cassette tape to record the remainder of the meeting. Participants tallied their positive statements, recorded the number of positive statements on a graph drawn on a sheet of quadrille paper (see Appendix C for Teacher 4’s graph), and then set goals for their next observation, which they indicated on the graph. The participants set their own goals for the next self-scoring session. Teacher 1 set goals to increase by 3-5 statements each session; Teacher 2, by 2-3 statements each session; Teacher 3, by 1-3 statements each session; Teacher 4, by 1-2 statements each session. Participants were not required to increase their goal the first time it was met, but were encouraged to do so if the goal had been met on two consecutive sessions.

The experimenter praised and encouraged the participants with respect to their performance, and provided suggestions, if requested, for meeting their goals. Questions regarding the participants’ impressions of the sample they had scored were included on the bottom of the participant data sheet; however review of these questions was not required. The experimenter also asked the participants what strategies, if any, they had used during the sample they had just scored. No specific interventions were implemented in this phase. See Appendix D for sample transcripts of the self-scoring meetings.

*Self-Scoring plus MotivAider®*

The self-scoring plus MotivAider® (an individually determined intervention) condition was only implemented with Teacher 3. In this condition Teacher 3 was also audiotaped daily and videotaped weekly. The self-scoring procedures from the previous
condition were continued, and meetings were again scheduled to be held four times per week and reduced by one per week. The experimenter provided a description of four possible interventions (tactile cueing, audio cueing, counter prompting, and self-recording) and the option of designing an individualized strategy (see Appendix E). The choice of interventions was made jointly; Teacher 3 chose a tactile cueing device called a MotivAider®, which resembles a pager (see Appendix F for a picture of the MotivAider®). Teacher 3 clipped the MotivAider® to an article of clothing for each recording session. The MotivAider® was programmed to vibrate on a variable time schedule, approximately once every 25 seconds. When the MotivAider® vibrated, Teacher 3 was to scan the classroom to find a student who was behaving appropriately and make a positive comment about that student’s behavior.

**Self-Scoring of Generalization Tapes**

Teacher 1 was the only participant included in this phase. Meetings took place once per week during this phase of intervention. These meetings were the same as the meetings in the initial self-scoring condition, with the exception that instead of scoring 5 minutes of an audiotape, Teacher 1 scored 5 minutes (of her choosing) from the generalization videotape. She continued to graph her number and rate of positive statements and to set goals for subsequent generalization observations. Daily audiotaping and weekly videotaping continued during this condition.

**Self-Scoring of Generalization Tapes; No Microphone**

This condition was also only implemented with Teacher 1, and it is identical to the previous condition with the exception that daily recordings during the regular recording
time were made (with Teacher 1’s knowledge) from the wall mounted video camera, instead of from the wireless microphone.

_Videotape Probes_

Teacher 1, Teacher 2, and Teacher 3 participated in this phase of the study. The wall mounted cameras were used for data collection during the last week of the study, which coincided with school wide testing. Testing sessions occurred each morning all week. Teacher 2 and Teacher 3 also discontinued use of the wireless microphone (Teacher 1 had discontinued use of the microphone in the previous condition). Participants were asked to specify several time periods each afternoon during which they would be engaged in whole-class instruction. The experimenter selected regular recording times selected from the times indicated by the participants. During this week no self-scoring meetings were held and participants were not informed of their use of positive or negative statements. Teacher 3 wore her MotivAider® during each day of this phase of the study.
CHAPTER 4

RESULTS

This chapter reports the results from the study. First, interobserver agreement results are presented with regard to occurrence measures of generic positive, specific positive, and negative statements. Second, treatment integrity data are reported. Next, data collected during baseline, self-scoring, and generalization conditions will be presented for each participant. Finally, the results of the social validity questionnaires given to participants and to current and future teachers are presented.

Interobserver Agreement

Two undergraduate and one graduate student collected interobserver agreement data on 40.2% of sessions. All sessions with IOA included agreement data on specific and generic positive statements and negative statements. Total (whole session) IOA ranged from 72.7% to 100% with a mean of 91.8%.

IOA data on generic positive statements ranged from 75% to 100% with a mean of 93%. IOA on specific positive statements ranged from 50% to 100% with a mean of 89.4%. IOA on negative statements ranged from 50% to 100% with a mean of 87.5%. See Table 4.1 for complete IOA results.
<table>
<thead>
<tr>
<th></th>
<th>Generic Positive</th>
<th>Specific Positive</th>
<th>Negative</th>
<th>Gen. Sessions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>92.9</td>
<td>86.9</td>
<td>84.8</td>
<td>97.2 (5)</td>
<td>91.2 (22)</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>92.2</td>
<td>93</td>
<td>92.6</td>
<td>93 (5)</td>
<td>92.5 (21)</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>92.7</td>
<td>92</td>
<td>85</td>
<td>94.4 (6)</td>
<td>91.8 (21)</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>96.6</td>
<td>100</td>
<td>87.5</td>
<td>96.1 (8)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>93</td>
<td>89.4</td>
<td>87.5</td>
<td>95.4 (16)</td>
<td>91.8 (72)</td>
</tr>
</tbody>
</table>

Table 4.1. Percent interobserver agreement across dependent variables and for generalization (Gen.) sessions for each participant. Numbers in parentheses indicate number of sessions scored for IOA.

**Treatment Integrity**

Analysis of treatment integrity data shows that with the exception of Teacher 3, 100% of target scoring sessions were held. Due to changes and disruptions in Teacher 3’s schedule, only three meetings were held during the first week of self-scoring, and only two meetings were held during the first week of self-scoring plus the MotivAider®. Therefore, instead of meeting eight times across these two phases, meetings were held only five times, producing a percent of target meetings score of 62.5%. Total percent of target meetings held (across all four participants) was 91.7%
For every self-scoring meeting that was held, all three elements of scoring (recording and setting a goal) were evident on participants’ data sheets and graphs. All four participants therefore had a score of 100% accuracy for the elements of each meeting.

Scheduled checks on data recording and supplemental materials verified that data were collected from the wall-mounted video cameras for scheduled generalization sessions and sessions after which the use of the wireless microphone had been discontinued. In addition, Teacher 3 wore and used her MotivAider® during each session in which its use was prescribed; visual checks were performed daily by the experimenter.

Positive and Negative Statements

Results for total positive and specific positive statements per minute by all participants are shown in Figures 4.1-4.4 (individual participant graphs, upper tier), and 4.6 (multiple baseline across participants graph). Total negative statements per minute by all participants are show in Figures 4.1-4.4 (individual participant graphs, lower tier), and 4.7 (multiple baseline across participants graph). Percent repeated statements are shown in Figure 4.5. Means for generic positive, specific positive, and negative statements per minute by all participants across conditions are shown in Figure 4.8.

Teacher 1

Baseline. The upper graph in Figure 4.1 shows total and specific positive statements per minute by Teacher 1 across all phases. Baseline data demonstrate a decreasing trend in total positive statements and a steady, low rate of specific positive statements. Training was held after session 12. Total positive statements initially increased, however an increasing trend was not evident overall after training.
Figure 4.1: Positive (top tier) and negative (bottom tier) statements per minute by Teacher 1 across conditions. The asterisk (*) indicates the occurrence of the training session, and the hatch marks (//) on the horizontal axes indicate a week-long school vacation.
The lower graph in Figure 4.1 shows total negative statements per minute by Teacher 1 across baseline, self-scoring, self-scoring of generalization tapes, and self-scoring of generalization tapes; no microphone. Baseline data demonstrate a variable rate of negative statements, with a decreasing trend prior to the training session and an increasing trend thereafter; the rate ranged from .1 to .7, with a mean of .3.

*Self-scoring.* Self-scoring was initiated after session 18. Within two sessions both total positive and specific positive rates exceeded baseline rates. Teacher 1’s rate of total positive statements steadily increased from 1.6 per minute to a high of 4.6, and her rate of specific statements increased from .2 to 1.8. After four weeks of self-scoring Teacher 1’s rate of total positive statements was 4.3 and specific positive statements were occurring at a rate of 1.1. Teacher 1 maintained these rates following a week-long school vacation that occurred after her 33rd session (indicated on Figure 1 by the hatch marks between school day 35 and school day 36). Teacher 1 always met her goal number of positive statements per 5-minute sample within two sessions.

Negative statements decreased during this phase from an initial high of .4 to 0 during the last two sessions.

*Self-scoring of generalization tapes.* After completing four weeks of self-scoring meetings Teacher 1 began self-scoring her weekly generalization tapes. Self-scoring was no longer conducted for daily audiotaped sessions. Teacher 1’s daily audiotaped rates of total positive and specific positive statements remained high, with a range of 4.1 to 5.1 and .9 to 1.9, respectively.
<table>
<thead>
<tr>
<th></th>
<th>Generic Positive</th>
<th></th>
<th>Specific Positive</th>
<th></th>
<th>Negative</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>17.0</td>
<td>2-34</td>
<td>2.3</td>
<td>0-7</td>
<td>4.6</td>
<td>1-11</td>
</tr>
<tr>
<td>Rate</td>
<td>1.1</td>
<td>.1-2.3</td>
<td>.2</td>
<td>0-.5</td>
<td>.3</td>
<td>.1-.7</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>13.8</td>
<td>6-18</td>
<td>1.5</td>
<td>0-3</td>
<td>6.5</td>
<td>4-10</td>
</tr>
<tr>
<td>Rate</td>
<td>.9</td>
<td>.4-1.2</td>
<td>.1</td>
<td>0-.2</td>
<td>.4</td>
<td>.3-.7</td>
</tr>
<tr>
<td>Self-Scoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>45.4</td>
<td>41-50</td>
<td>20</td>
<td>13-29</td>
<td>.1</td>
<td>0-1</td>
</tr>
<tr>
<td>Rate</td>
<td>3.0</td>
<td>2.7-3.3</td>
<td>1.3</td>
<td>.9-1.9</td>
<td>0</td>
<td>0-.1</td>
</tr>
<tr>
<td>Self-Scoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalization; No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microphone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>46.6</td>
<td>35-58</td>
<td>15.6</td>
<td>56-71</td>
<td>.3</td>
<td>0-1</td>
</tr>
<tr>
<td>Rate</td>
<td>3.2</td>
<td>2.3-4.3</td>
<td>1.1</td>
<td>.8-1.5</td>
<td>0</td>
<td>0-.1</td>
</tr>
<tr>
<td>Videotape Probes a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>25</td>
<td>n/a</td>
<td>7</td>
<td>n/a</td>
<td>1</td>
<td>n/a</td>
</tr>
<tr>
<td>Rate</td>
<td>1.7</td>
<td>n/a</td>
<td>.5</td>
<td>n/a</td>
<td>.1</td>
<td>n/a</td>
</tr>
</tbody>
</table>

*a – one session conducted in this condition

Table 4.2. Means and ranges of dependent variables calculated as number of statements and rate of statements for Teacher 1.
Negative statements continued to occur at low rates during this condition. With the exception of the first data point (.1) no negative statements occurred during this condition.

*Self-scoring of generalization tapes; no microphone.* After seven sessions of self-scoring generalization tapes, the use of the microphone was discontinued with Teacher 1. Self-scoring of generalization tapes continued through this condition. Daily sessions continued to demonstrate a high, although moderately decreasing trend for total positive statements, and an increasing trend for specific positive statements. The rate of total positive statements ranged from 3.7 to 4.8, and the rate of specific positive statements ranged from .8 to 1.5.

*Repeated positive statements.* Figure 4.5 (see page 79) shows the percentage of positive statements that were repeated. Teacher 1’s data demonstrate a slight decrease in repeated total positive statements and a slight increase in repeated specific positive statements.

*Generalization data.* The range of total positive statements recorded during generalization sessions prior to self-scoring generalization tapes was .3 to .6, with a mean of .4 positive statements per minute. Self-scoring generalization tapes initially produced an increase in the next session (1.2 positive statements per minute), however the next two generalization data points were comparable with pre-intervention data. The fourth self-scoring generalization data point increased to 1.3, and the final point, collected during the last phase in which all self-scoring had been discontinued, was 2.1.
Figure 4.2: Positive (top tier) and negative (bottom tier) statements per minute by Teacher 2 across conditions. The asterisk (*) indicates the occurrence of the training session, and the hatch marks (//}} on the horizontal axes indicate a week-long school vacation.
Teacher 2

Baseline. Teacher 2’s data are shown in Figure 4.2. Teacher 2’s baseline data show a gradually decreasing trend. Prior to training the range of total positive statements per minute was 0 to .9 with a mean of .41, and after training the range was .1 to .7 with a mean of .39. The range of specific positive statements per minute prior to training was approximately the same: .1 to .9, with a mean of .41. The range of specific positive statements after training was .1 to .8, with a mean of .39. The second tier of Figure 4.2 shows the negative statements per minute for Teacher 2. These data were variable, with a range of 0 to .5 per session.

<table>
<thead>
<tr>
<th></th>
<th>Generic Positive</th>
<th>Specific Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>6.2</td>
<td>0-14</td>
<td>5.8</td>
</tr>
<tr>
<td>Rate</td>
<td>.4</td>
<td>0-.9</td>
<td>.4</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>6</td>
<td>2-11</td>
<td>5.7</td>
</tr>
<tr>
<td>Rate</td>
<td>.4</td>
<td>.1-.7</td>
<td>.4</td>
</tr>
<tr>
<td>Self-Scoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>29.6</td>
<td>6-53</td>
<td>11.9</td>
</tr>
<tr>
<td>Rate</td>
<td>2</td>
<td>.4-3.5</td>
<td>.8</td>
</tr>
<tr>
<td>Videotape Probes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>6.3</td>
<td>0-12</td>
<td>.3</td>
</tr>
<tr>
<td>Rate</td>
<td>.4</td>
<td>0-.8</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.3. Means and ranges of dependent variables calculated as number of statements and rate of statements for Teacher 2.
**Self-scoring.** Teacher 2’s total positive and specific positive statements per minute gradually increased over the 21 sessions of self-scoring. After an initial low rate of .8 total positive statements per minute, data were variable, but ascending, to a final high of 4.7. Due to the onset of school-wide testing, self-scoring of generalization tapes was not implemented with Teacher 2. Specific positive statement data demonstrate an increasing trend after self-scoring was introduced. Teacher 2 always met her goal number of positive statements per 5-minute sample within two sessions.

During this condition the number of negative statements per minute ranged from 0 to .3; negative statements show a slight decrease across this phase. No negative statements were recorded in three of the last four sessions during the self-scoring condition.

**Repeated positive statements.** Figure 4.5 shows Teacher 2’s percent of positive statements that were repeated. Teacher 2’s data demonstrate an increase in both total and specific repeated positive statements during the self-scoring condition.

**Generalization data.** No self-scoring was conducted for Teacher 2’s generalization tapes, however generalization data show an increasing trend through the baseline and self-scoring conditions, from a low of .1 positive statements per minute during the second session (while baseline conditions were in effect) to a high of 1.4 in the seventh generalization session during the self-scoring condition. Generalization data during the final week of the study demonstrated low variable rates of positive statements with a range of 0 to .9 and a mean of .5. During this final condition negative statements were stable at a rate of .1 per session.
Figure 4.3: Positive (top tier) and negative (bottom tier) statements per minute by Teacher 3 across conditions. The asterisk (*) indicates the occurrence of the training session, and the hatch marks (//) on the horizontal axes indicate a week-long school vacation.
Teacher 3

**Baseline.** Teacher 3’s data are shown in Figure 4.3 Baseline data for Teacher 3 demonstrate low variable rates of total positive and specific positive statements. Total positive statements ranged from .1 to 1 per minute, and specific positive statements ranged from 0 to .5. Teacher 3’s highest baseline data points occurred after her training session, which was before her 23rd session and immediately after the school vacation week. Her data after the training session was similar to Teacher 1’s and Teacher 2’s data in that there was a moderate, but short-lived increase.

Negative statements for Teacher 3 were variable during baseline sessions. They ranged from 0 to .6. This high rate occurred after Teacher 3’s training session.

**Self-scoring.** Self-scoring was only implemented for Teacher 3 for four school days because recording sessions were suspended due to her attendance at daily curriculum training meetings. The rates of positive statements in this phase were similar to the rates obtained after the training session; there was not a substantial increase. Teacher 3 met her goal number of positive statements per 5-minute sample within three sessions.

Negative statements occurred at lower rates during this condition than while baseline conditions were in effect. One negative statement was recorded in both the first and third sessions of this condition, and three negative statements were recorded in both the second and fourth sessions. The mean rate of negative statements during self-scoring was .13.
Self-scoring plus MotivAider®. After Teacher 3 completed her curriculum training and daily recording sessions resumed, the experimenter asked her to select an intervention to increase her rate of positive statements from the menu of intervention choices found in Appendix E. She and the experimenter mutually agreed on using the MotivAider®. This was a small pager-like device that Teacher 3 clipped to her clothing. It was programmed to vibrate on a variable-time 25 (VT 25) seconds schedule. This generated a prompt for Teacher 3 to say something positive to her students at a rate of approximately 2.4 times per minute. Data demonstrate that during the first two sessions of this condition Teacher 3’s rate of positive statements did increase to a rate of 2.0 and 2.3, however this was followed by a decrease to 1.3 in the fourth and final session in this condition. Mean rates of positive statements were higher in this condition than in the two previous conditions, however these rates were not maintained.

Negative statements continued to decrease during this condition. A total of three negative statements were recorded (one in each of the first three sessions) during the self-scoring plus MotivAider® phase.

Repeated positive statements. Figure 4.5 shows Teacher 3’s percent of positive statements that were repeated. Teacher 3’s data do not show a consistent trend during the course of the study.
<table>
<thead>
<tr>
<th></th>
<th>Generic Positive</th>
<th>Specific Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>4.7</td>
<td>0-11</td>
<td>1.7</td>
</tr>
<tr>
<td>Rate</td>
<td>.3</td>
<td>.1-.7</td>
<td>.1</td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>7.8</td>
<td>3-11</td>
<td>2.4</td>
</tr>
<tr>
<td>Rate</td>
<td>.5</td>
<td>.2-.7</td>
<td>.2</td>
</tr>
<tr>
<td><strong>Self-Scoring</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>9</td>
<td>5-15</td>
<td>4.5</td>
</tr>
<tr>
<td>Rate</td>
<td>.6</td>
<td>.3-1</td>
<td>.3</td>
</tr>
<tr>
<td><strong>Self-Scoring plus MotivAider®</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>20</td>
<td>14-26</td>
<td>7.5</td>
</tr>
<tr>
<td>Rate</td>
<td>1.3</td>
<td>.9-1.7</td>
<td>.5</td>
</tr>
<tr>
<td><strong>Videotape Probes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>9</td>
<td>5-13</td>
<td>2</td>
</tr>
<tr>
<td>Rate</td>
<td>.6</td>
<td>.3-.9</td>
<td>.1</td>
</tr>
</tbody>
</table>

Table 4.4. Means and ranges of dependent variables calculated as number of statements and rate of statements for Teacher 3.

**Generalization data.** Teacher 3’s generalization data demonstrate low rates across all sessions, with an increase (relative to previous conditions) during the final phase of the study (after self-scoring and daily audiotape recording ended). Teacher 3 did continue to wear the MotivAider® during this phase of the study. Rates of negative statements remained at zero. Self-scoring of generalization tapes was not initiated with Teacher 3 due to time constraints and the onset of school wide testing.

**Teacher 4**

*Baseline.* Baseline data for Teacher 4 are shown in Figure 4.4, and include six weekly videotaped sessions and nine daily audiotaped sessions. Baseline data
Figure 4.4: Positive (top tier) and negative (bottom tier) statements per minute by Teacher 4 across conditions. Hatch marks (//) on the horizontal axes indicate a week-long school vacation.
demonstrate low variable rates of positive statements. Rates ranged from .3 to 1.2 total positive statements and 0 to .4 specific positive statements. Training was not conducted with Teacher 4 due to the consistent and inconsequential effects of training with the other three participants; in addition, only one week of the school year was available to implement the self-scoring condition.

Teacher 4’s rates of negative statements ranged from 0 to .7 during baseline, with a mean of .1. Rates were at or below .3 during baseline, with the exception of the fourth audiotaped session which had a rate of .7 negative statements.

<table>
<thead>
<tr>
<th></th>
<th>Generic Positive</th>
<th>Specific Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>7.6</td>
<td>4-18</td>
<td>.9</td>
</tr>
<tr>
<td>Rate</td>
<td>.5</td>
<td>.3-1.2</td>
<td>.1</td>
</tr>
<tr>
<td>Self-Scoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>18.2</td>
<td>8-23</td>
<td>6.4</td>
</tr>
<tr>
<td>Rate</td>
<td>1.6</td>
<td>.8-1.9</td>
<td>.6</td>
</tr>
</tbody>
</table>

Table 4.5. Means and ranges of dependent variables calculated as number of statements and rate of statements for Teacher 4.

**Self-scoring.** Teacher 4’s self-scoring data demonstrate an increase in total and specific positive statements per minute. Although this condition was brief, the mean rates for total and specific positive statements were more than 3 times the mean rates during
baseline. Teacher 4 met each goal number of positive statements per 5-minute sample within one session.

Teacher 4’s negative statements decreased during this condition. On four of the five days during which recording took place, Teacher 4 made no negative statements to his students.

Repeated positive statements. Figure 4.5 shows Teacher 4’s percent of positive statements that were repeated. Teacher 4’s data do not show a consistent trend during the course of the study.
Figure 4.5: Percentage of positive statements repeated by Teacher 1 (T1), Teacher 2 (T2), Teacher 3 (T3), and Teacher 4 (T4). Asterisks (*) indicate training sessions and hatch marks (///) on the horizontal axes indicate a week-long school vacation.
Group Summary

Figure 4.6 shows the rate of specific positive and total positive statements by all four participants across phases. Data demonstrate descending trends or low stable data paths for all participants during baseline, and three of four participants’ data show increases in positive statements when (and only when) the self-scoring condition was implemented, which lends weight to the experimental control of the procedures and illustrates the experimental logic of the multiple baseline design. In addition, Teachers 1, 2, and 3 all show small, but short-lived increases in their total positive statements after their training sessions.

Figure 4.7 shows the rate of negative statements by all four participants across phases. Changes in this dependent variable are also evident only when self-scoring is implemented with each participant, and the rate of negative statements decreased consistently after self-scoring was implemented.

Figure 4.8 shows the mean rate of generic positive, specific positive, and negative statements per minute for all participants across conditions. This figure demonstrates consistent increases in both categories of positive statements during and after the implementation of self-scoring, for all participants. In addition, all participants’ data demonstrate decreases in the rate of negative statements from baseline to self-scoring and subsequent phases.
Figure 4.6: Number of total and behavior-specific positive statements per minute by Teacher 1 (T1), Teacher 2 (T2), Teacher 3 (T3), and Teacher 4 (T4). Asterisks (*) indicate training sessions and hatch marks (//) on the horizontal axes indicate a week-long school vacation.
Figure 4.7: Number of negative statements per minute by Teacher 1 (T1), Teacher 2 (T2), Teacher 3 (T3), and Teacher 4 (T4). Asterisks (*) indicate training sessions and hatch marks (///) on the horizontal axes indicate a week-long school vacation.
Figure 4.8: Mean number of statements per minute by all participants across conditions. Self scoring is abbreviated “SS,” generalization is abbreviated “Gen.,” and “MA” refers to the intervention selected by Teacher 3, the MotivAider®.
Social Validity

Participants’ Opinions

Results of the social validity questionnaire that was given to participants (see Appendix G) are shown in Table 4.6. Overall, participants rated the intervention highly acceptable both before and after intervention. All participants, with the exception of Teacher 3, gave the intervention a rating of at least 6 (out of 7) prior to starting the intervention. Teacher 3 had the greatest change in ratings from baseline to intervention; her score increased 1.5 points to 6.8. Teacher 2 gave the intervention a 7 across all questions both before and after the intervention.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Baseline</th>
<th>Post-Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>Teacher 1</td>
<td>6.6</td>
<td>4-7</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>7.0</td>
<td>n/a</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>5.3</td>
<td>3-7</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>6.1</td>
<td>5-7</td>
</tr>
</tbody>
</table>

Table 4.6. Treatment acceptability rating scores by participants, where 1 = not acceptable and 7 = highly acceptable.

Current and Future Teachers’ Opinions

Table 4.7 summarizes the responses to the social validity questionnaire that was given to current and future teachers (see Appendix H). This group comprised 26
undergraduate students and 18 part time graduate students in special education.

Responses demonstrate higher ratings for intervention video samples than for baseline video samples across all four participants.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th></th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>Teacher 1</td>
<td>4.5</td>
<td>4.2-5.0</td>
<td>4.6</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>3.3</td>
<td>3.1-3.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>4.4</td>
<td>4.2-4.6</td>
<td>6.0</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>4.0</td>
<td>3.7-4.2</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Table 4.7. Social validity survey means (across all responses) and ranges (across questions) for all participants, where 1 = very poor and 7 = very good.
CHAPTER 5

DISCUSSION

This chapter addresses the results of the study investigating the effects of self-scoring on teachers’ positive and negative statements during classroom instruction. Limitations of the study, results relative to the research questions, implications for practice, and suggestions for future research are discussed.

Limitations of the Study

Participant Characteristics

All four teachers were initially identified for participation in this study because they had requested assistance with classroom management from the school’s director or from the experimenter. Upon being promoted to a teaching position, Teacher 1 had told the experimenter that any help would be appreciated, and that she was “up for anything.” Teacher 2 identified herself as a potential participant two months prior to the start of data collection when she asked the experimenter how to obtain a MotivAider®. Teacher 3 had previously referred a student to the school’s Intervention Assistance Team (IAT), of which the experimenter was a member. Teacher 3 had also indicated that she was willing to try any and all suggestions for improving social and academic behavior in her class.
Teacher 4 had requested assistance with his second grade classes in September, almost five months before consent forms for the study had been distributed. The fifth participant, who was eventually excluded from the study, had also requested assistance with classroom management from the experimenter over the course of the previous school year. It is possible that these teachers’ interest in improving their classroom management strategies and their willingness to change their behavior makes the results of this study less likely to generalize to teachers who are not interested in changing their own behavior. In total, eight informational letters and consent forms were distributed to teachers at the school, but only five of these eight teachers returned signed forms. The other three teachers declined to participate, saying that they thought the study would be too disruptive or that they had participated in videotaped self evaluation during their student teaching and were not interested in doing so again. Such responses indicate that there are certain prerequisite characteristics that make the success of this intervention more likely.

Length of the Study

A second limitation concerns the length of the study. This study was conducted over a 14-week period, beginning in early February; the study terminated a week before Memorial Day. Two participants (Teacher 1 and Teacher 2) were able to complete a full course of self-scoring. That is, they began the self-scoring intervention with four meetings per week and ultimately participated in four weeks of self-scoring so that they reduced their meetings to once per week. Teacher 3 was not able to participate for more than one week in each self-scoring condition, and was not able to meet for four times during either week. Teacher 4 was able to meet four times, but only for one week before
school-wide testing began. During debriefing interviews at the conclusion of the study, the consensus among the participants was that they would have preferred the intervention to have started earlier in the school year. This would have allowed for more time for self-scoring as well as generalization of their new skills in positively managing their classrooms to other times of the day. In addition, they could have implemented their own self-scoring strategies after their meetings with the experimenter were faded out.

*Treatment Integrity*

Treatment integrity data for Teacher 3 revealed that this participant only attended 62.5% of the meetings that were targeted for her two intervention phases. Although all components of each meeting that was held were implemented accurately, the omission of several prescribed meetings could potentially account for the ineffectiveness of the independent variable relative to her small increase in rate of positive statements. Teacher 3 did score three sessions during the first self-scoring condition and two sessions during the self-scoring plus MotivAider® condition, however in missing the opportunity to score her own behavior three times, it is possible that whatever variables that were responsible for changes in the other participants’ behavior were not operating in the case of Teacher 3.

The optimum number of meetings was not assessed in this study. Future research could investigate whether four meetings per week are necessary, or whether perhaps a constant number of meetings per week, for example, two, would be adequate to produce changes in the rates of teachers’ positive statements. It is not clear that holding fewer than the target number of meetings per intervention condition was the sole variable
responsible for the lack of change in Teacher 3’s data, however the results of the study should not be considered without taking this information into account.

Data on Student Behavior

A fourth potential limitation of this study is that no data were collected on student behavior or work products. The four teachers who participated in this study reported improvements in student behavior. Teacher 1 noted that students had learned to raise their hands quietly instead of shouting out answers or saying “Oooh!” while raising their hands. Teacher 2 noted that students did a better job of working independently and quietly, and also of working appropriately in small groups when she gave them that option. She also said that many students had become more efficient at transitioning between activities during independent work time. Teacher 3 noted that when she praised specific students for behavior such as being prepared for lessons or sitting up straight, other students followed suit, treating her positive statement like an instruction. Teacher 4 was so impressed by his students’ behavior that he routinely walked his classes down the hall to show them off to other teachers. These anecdotal reports add weight to the effectiveness of this intervention, but they are not sufficient evidence of improvement in student behavior.

Although behavioral and other special education research over the last 40 years has demonstrated that increases in positive teacher attention result in increases and improvements in student academic and managerial behavior, the lack of data on student behavior in this study leaves doubts about the social significance of the results. In other words, without knowing that student behavior did, in fact, improve when teacher positive
statements increased, it is difficult to draw conclusions about the importance of this study. Participants indicated that they found the intervention acceptable, and even helpful, and current and future teachers noted improvements in teachers’ behavior across the course of the study, however no assessment was conducted of the most significant and socially important outcome of the study: improvement in student learning. If data on student behavior during this study were to demonstrate that student academic and social behavior actually declined while teacher positive statements increased, the results of the study and their implications for practice would be questionable.

Definitions

A fifth limitation to the study concerns the categories and definitions of positive and negative behavior that were developed and used for scoring. Specifically, no non-vocal behaviors, such as facial expressions, gestures, and physical contact were incorporated into the definitions. Although Kazdin and Klock (1973) did demonstrate that non-vocal behavior can improve student behavior, research has yet to demonstrate that either vocal or non-vocal behaviors are more desirable for use in changing student behavior.

The addition of teacher facial expressions as positive and negative teacher behaviors could have changed the results of the study. For example, if negative facial expressions (e.g., scowls) were included in the definitions, it is possible that these would have increased the number of negative behaviors observed during each session. Similarly, adding positive facial expressions and non-verbal behavior such as a high five or a pat on the back could have increased the number of positive behaviors recorded for each teacher.
This change in the definitions would have also caused a change in the recording system: in order to capture teacher non-vocal behaviors, all sessions would have to have been videotaped instead of audiotaped.

Research Questions

Research Question 1: What are the effects of self-scoring the number of positive statements from audio and videotapes of instruction on teachers’ rates of positive statements delivered during subsequent instruction?

Data from the self-scoring conditions implemented with Teacher 1, Teacher 2, and Teacher 4 suggest that this intervention resulted in an increase in positive statements during subsequent instructional sessions that were similar to the scored sessions. That is, daily audiotaping data revealed higher rates of positive statements; however rates of positive statements were only slightly higher during generalization sessions for Teacher 1, Teacher 2, and Teacher 3. Teacher 3’s daily audiotaping data did not demonstrate a substantial increase over baseline rates during the self-scoring phase, however Teacher 3 also did not participate fully in this condition with respect to participating in the prescribed number of scoring sessions.

It is difficult to determine what consequences specifically functioned to reinforce teachers’ use of increasing numbers of positive statements, since no immediate consequences for each positive statement were discernible to the experimenter or to the teachers themselves. One possibility is that the sound produced and heard when delivering a positive statement to a student was reinforcing in itself: hearing oneself praising a student was automatically reinforcing, even if the student’s behavior did not
immediately improve. Over time student behavior did improve, however any change in student behavior would have been too far removed from teachers’ positive statements to be considered a reinforcer for those statements.

Another possibility is that listening to the audiotapes, recording and graphing positive statements, and setting goal rates for the next scoring meeting resulted in participating teachers forming a rule with respect to their use of positive statements during subsequent observations. Rules functioned as indirect acting contingencies, in that they could control behavior despite the apparent delay between increases in positive statements and subsequent improvements in student behavior (Malott, Whaley, & Malott, 1991). It is possible that the rule itself established an aversive condition for the teachers in that each teacher felt pressure to meet their individual goals. Teacher 1 commented that she preferred to set her goals low, rather than high, because she would be disappointed if she failed to meet her goals. Each positive statement emitted may have alleviated the pressure to meet the goal, therefore reinforcing positive statements by reducing the aversive condition established by the rule.

Research Question 2: What are the effects of an intervention package consisting of self-scoring the number of positive statements from audio and videotapes of instruction and individually determined interventions on teachers’ rates of positive statements delivered during subsequent instruction?

Teacher 3 was the only participant to add an individually determined intervention, the MotivAider®, to the self-scoring procedure. Results demonstrated that while this initially coincided with an increase in positive statements, the increase was short-lived.
Teacher 3’s rate of positive statements quickly returned to its previous level. Given the lack of treatment integrity for Teacher 3, however, it is not surprising that this intervention was not effective. Simply put, this condition was not implemented as planned, and so determining its effects poses a challenge. In addition to the lack of treatment fidelity, it is also interesting to note that Teacher 3 gave the self-scoring intervention the lowest pre-intervention acceptability score on the TARF-R, across all four participants. It is also possible that whatever variables contributed to her lower than average treatment acceptability score also contributed to her low rates of positive statements during this phase of the study.

Research Question 3: What are the effects of self-scoring the number of positive statements from audio and videotapes of instruction on teachers’ rates of specific and varied positive statements, and negative statements, delivered during subsequent instruction?

Teacher 1 and Teacher 2’s data demonstrate the most dramatic increases in specific statements across the four participants. Teacher 4’s data show a small increase as well. Teacher 3’s data do not show an increase above baseline levels during self-scoring.

One reason that specific statements may have increased, despite the fact that they were not scored separately by the participants, is the nature of the data sheet that was used. Because new statements were written on a new line on the data sheet, as opposed to repeated statements being recorded by their respective time indexes, filling out the data sheet provided the participants with a visual display of the variety of their positive statements. Teacher 1 did comment that her goal was to fill up the data sheet. While she
could have accomplished this simply by using a variety of generic statements, she aimed to increase her behavior-specific statements as well. In comparing her rate of specific statements to her percent of repeated specific statements (see Figures 4.1 and 4.5), it is important to note that the variability in repeated specific statements during baseline is probably due to the low rate of specific statements overall. If any participant made one or zero specific positive statements, the percent of repeated specific positive statements would be zero. Teacher 1’s repeated specific positive statements were highly variable during baseline while rates of specific positive statements were low. Teacher 1’s data do not show a substantial decrease in percent of repeated specific positive statements; however her percent of repeated specific positive statements did decrease and become less variable from baseline to self-scoring and self-scoring of generalization videotapes.

Data on repeated statements are variable for the other three participants. Teacher 2’s data seem to increase across her self-scoring condition, however they are variable throughout her baseline phase, again, probably because of low rates of specific positive statements during this condition. Teacher 1’s data show a slight decrease in total repeated positive statements across all phases however this decrease is probably not meaningful, in the sense that it was unlikely to be detected by her students or other observers. Teacher 3’s and Teacher 4’s data on repeated statements were variable throughout the study.

The most consistent effect of the self-scoring intervention was a reduction in negative statements. All four participants demonstrated decreases in their use of negative statements after the self-scoring intervention was implemented, and of the three
participants who participated in data collection after self-scoring was discontinued, all three continued to demonstrate a reduction in negative statements.

The reduction in negative statements is an interesting effect of the intervention, given that negative statements were not recorded during self-scoring sessions. It is possible that the sound of their negative statements on the tape was aversive to the participants, and in the future they aimed to make fewer negative statements so as to avoid having to listen to them. Hearing the negative statements on the tape may also have made negative statements more salient to the participants, so that they were more aware of them during teaching.

Another possibility is that in their attempts to increase their positive statements, teachers began to use one student’s misbehavior as an opportunity to praise a student who was behaving appropriately. For example, if one student was out of his seat without permission, a teacher might say, “Thank you, Carrie, for staying in your seat.” This specific positive statement could potentially serve as reinforcement for Carrie’s staying in her seat, as well as an instruction to the student who was out of his seat. In addition, using positive statements in this way helped the participating teachers find an alternative to using negative statements to change undesirable student behavior.

Sutherland and Wehby (2001) also noted a decrease in negative statements (reprimands) in their study, despite the fact that they also did not include self-scoring negative statements in their treatment. They hypothesized that the increases in students’ correct responses that resulted from increases in rates of praise and opportunities to respond produced a decrease in student misbehavior, thus reducing the need for negative
statements. While plausible, it is not possible to evaluate this conclusion without data on student misbehavior.

*Research Question 4: What are the effects of an intervention package consisting of self-scoring the number of positive statements from audio and videotapes of instruction and individually determined interventions on teachers’ rates of specific and varied positive statements, and negative statements, delivered during subsequent instruction?*

Teacher 3 was the only participant to be included in this condition, which combined self-scoring and use of the MotivAider®. The most noticeable change in Teacher 3’s data during this phase was for her use of negative statements. Despite the fact that her positive statements showed a short-lived increase, her negative statements were very low (a range of 0-.1 per minute) during the four days of this condition. The combination of the MotivAider® and self-scoring did not have a noticeable effect on specific positive statements or on the percentage of total repeated positive statements. Specific positive statements were repeated slightly more often during the self-scoring plus MotivAider® phase, however the fourth and final data point in this condition is 0%, which is comparable to results in previous phases.

*Research Question 5: What are teachers’ perceptions about the intrusiveness and effectiveness of self-scoring the number of positive statements from audio and videotapes of instruction and/or an intervention package consisting of self-scoring the number of positive statements from audio and videotapes of instruction and individually determined interventions?*
The Treatment Acceptability Rating Form – Revised (TARF-R; Reimers & Wacker, 1988) results demonstrated varying increases in teachers’ acceptability scores from the pre-self-scoring to the post-intervention administrations of the assessment. All participants provided mean scores that were above the neutral (4) rating in the first assessment. Teacher 2 answered all questions with the highest possible rating on both administrations of the TARF-R. Teacher 3’s data demonstrate the biggest increase from the first administration to the second, which is interesting in light of the fact that she was the only participant to add an individually determined intervention to the self-scoring procedure, and the failure of the self-scoring intervention by itself to produce a significant increase in her rate of positive statements. Teacher 3 reported that the intervention did make her more aware of appropriate student behavior, but that she was wary of delivering too much praise, for fear her students would find this behavior disingenuous.

Research Question 6: What are current and future teachers’ perceptions of “before intervention” and “after intervention” videotaped samples of participating teachers’ instruction?

Results of the social validity assessment of current and future teachers’ opinions demonstrated increases in scores from baseline to post-intervention for all participants; all participants’ post-intervention scores were above the neutral (4) mark on the rating scale. Teacher 3’s data are again of interest, as her post-intervention score is higher than the other participants’ post-intervention scores by more than a full point. These data were unexpected given that Teacher 3 had the lowest rate of positive statements of all four
participants. Despite the fact that each participant’s “best” videotaped session was used as the post-intervention video shown to the consumer groups, it is possible that there are qualitative differences in Teacher 3’s behavior in this particular video that were not addressed or measured by the dependent variables as they were defined for this study. Such qualitative differences require further scrutiny and should be assessed for possible inclusion in future investigations.

**Summary of Research Findings**

This study contributes to previous research on strategies for increasing teachers’ rates of positive, or praise statements by demonstrating the effectiveness of a self-scoring package intervention. This intervention comprised several elements of previous research, including feedback and goal setting (Martens et al., 1997; Sharpe et al., 2002) and self-scoring (Sutherland & Wehby, 2001).

This study confirms previous investigations on the effectiveness of self-scoring to teachers’ rates of praise (Cusimano, 1987; Sutherland & Wehby, 2001). While future research on self-scoring is needed, the consistency in the existing findings warrants discussion. Of particular interest is the reduction in negative statements recorded in the present study and in the investigation by Sutherland and Wehby. Participants in both studies demonstrated a decrease in a dependent variable that was not targeted for self-scoring, or even for discussion between the teacher and the experimenter. Further study is needed on this collateral behavior change.

This study extended previous research by incorporating an extended and gradually faded self-scoring phase. In contrast to Sutherland and Wehby’s (2001) study,
participating teachers met with the experimenter for each self-scoring session. This may have contributed to the more robust findings in the present study, however this also raises the question of how best to fade the experimenter’s presence from the self-scoring sessions. As effective as any strategy to change teacher behavior may be, it will best serve the teacher if it can be employed independently and does not create teacher dependence on a coach or experimenter.

This study also extended the research of Sutherland and Wehby (2001) by employing a single-subject, as opposed to a groups-comparison design. The use of a single subject design enabled continuous contact with the data on teacher behaviors, and allowed for implementation of changes in the independent variable as changes were observed in the dependent variable (e.g., Teacher 3’s self-selected intervention, the MotivAider®). In addition, variability in each teacher’s data enabled more precise analysis of the dependent variables. For example, Teacher 3’s data in the self-scoring plus MotivAider® condition demonstrate an increase, and then a rapid decrease in her rate of positive statements (see Figure 4.3). Her summary data across conditions, however, seem to indicate that with each successive phase her positive statements increased (see Figure 4.8). While her mean number of positive statements did increase in each condition, the data do not support the conclusion that self-scoring was effective with Teacher 3. A closer analysis of the session-by-session data of Sutherland and Wehby’s participants might yield more detailed information about the effects of self-scoring over time for individual teachers.
Although this study combined previous research on several behavior-change strategies, it is difficult to draw conclusions regarding the effects of these individual strategies on the changes in the behavior of participants in the present study. A component analysis is necessary to delineate the individually effective components, or to determine that the package alone was responsible for the effects observed. A component analysis is not required, however for any teacher attempting to improve his rate of positive statements; pre-service and in-service teachers are encouraged to employ the individual or combined strategies that fit best with their classrooms and their existing resources.

Implications for Practice

The results of this study suggest that self-scoring is a useful tactic that teachers can use to increase the positive statements they deliver to their students during classroom instruction. Based on data collected with Teacher 3, it appears that consistent application of a self-scoring intervention is important to its success; inconsistent meetings and scoring are not likely to yield dramatic results. Teachers and supervisors in applied settings may find that four meetings per week take up too much time, and that perhaps a smaller number of meetings would produce similar effects, so long as those meetings were held consistently.

Results of the study also suggest that when used in an applied setting – where experimental control is not as much a cause for concern as it is in empirical studies such as this one – it would perhaps benefit teachers and their supervisors to collect data samples at varied times and across instructional activities throughout the day for future
scoring, rather than adhering to one recording time only per day. Collecting audiotape or videotape samples throughout the day, and perhaps making the recording schedule unpredictable for the teachers, would promote generalization of teaching skills across subjects and students.

If a teacher is interested in implementing a self-scoring procedure for herself, the procedures followed in this study should be modified to suit the teacher’s needs and her resources. A formal data sheet may not be necessary; counting the total number of positive statements heard within 5 minutes might suffice. Also, a teacher should consider recording her instruction at varying times throughout the school day, instead of always recording at the same time; this will help to promote generalization. The teacher might also implement a public posting component, in which she would graph the total number of positive statements recorded each day, along with each day’s goal, to aid in the recruitment of feedback from co-workers.

A related implication of this study for applied settings concerns the cost of the intervention, in terms of both time and money. This experiment required daily data collection for four participants, and often three or four self-scoring meetings per participant per week. In an applied setting, the scheduling and frequency of recording and meetings should be modified to best fit a teacher’s schedule. Also, due to the requirements of the experimental design used and the need to demonstrate experimental control, several procedures were implemented in this study that increased the financial expense of the equipment required; these include the wall-mounted video cameras and recording equipment used for the generalization probes, and the digital video camera
used simultaneously once per week with the audio recording equipment. The combined cost of this equipment exceeded $600, in addition to the $300 spent on the audio recording equipment. Teachers who wish to implement a self-scoring procedure in their classrooms simply to increase their own rates of positive statements do not need such specialized equipment. A cassette recorder, placed near enough to the teacher to pick up his or her voice, is the only equipment required.

Another implication concerns teachers’ skill levels prior to implementing self-scoring. Self-scoring is a useful tactic for teachers who know what constitutes a positive statement and what does not. Teacher 1, Teacher 2, and Teacher 3 demonstrated agreement with the experimenter on what praise is and is not when they completed their training sessions. If they had not obtained high agreement scores when recording the positive statements of the non-participating teacher on the videotape, further training and scoring would have been conducted. Teachers who do not demonstrate an understanding of what behaviors (on their part) are likely to function as reinforcers for their students would benefit from additional training and support in scoring their own audio and videotapes. Such training could take the form of role playing with a supervisor and observing and recording the behavior of fellow teachers who do demonstrate high rates of positive statements and whose students demonstrate appropriate behavior in class.

A final, but no less important, implication of this study concerns the effects of the training session for Teacher 1, Teacher 2, and Teacher 3. This training session could be likened to a brief workshop on increasing positive statements: participants were shown the data sheet that they would be using, and were told that they would be observing a
teacher who used a high number of positive statements while teaching. They then watched the tape and recorded the positive statements made by this teacher. Similar activities are conducted during inservice trainings, and generally, they have the same effects: small and short-lived. Simply telling a teacher to use more positive statements, and even showing a teacher what positive statements are and that they can be incorporated into classroom instruction, does not seem to be effective in increasing that teacher’s own positive statements during his or her own instruction. A major benefit of the self-scoring procedure is that it makes the use of positive statements more salient, and in setting goals, gives the teacher a specific rule to follow.

Suggestions for Future Research

*Generalization*

Perhaps the most important issue that should be addressed in future research is generalization of increased rates of positive statements across the school day. In this study, generalization data were collected at times other than those targeted for daily data collection, however no strategies were implemented specifically to promote generalization of positive statements to times other than those scheduled for audio recording. Although Teacher 1 did score several generalization videotapes, it appears that scoring these tapes only once each week was not effective in increasing her rate of positive statements during the generalization data collection sessions.

One strategy for promoting generalization is to conduct multiple audio recording sessions daily, but to only use one session for self-scoring. The decision of which session to use would be made by the experimenter, or an indiscriminable contingency could be
implemented. An example of an indiscriminable contingency would be the experimenter randomly assigning numbers 1-n (n = the number of recordings per day) and having the teacher select one number and scoring 5 minutes of the selected tape. The teacher would know when recording was occurring, but would not know which tape would be scored.

Another tactic for increasing positive statements across the school day is conducting daily recording sessions on an unpredictable schedule with a wireless surveillance camera or another similarly covert device; these tapes could be used for the daily self-scoring sessions. It might be advantageous to first record a participating teacher once during a limited predetermined time period, for example, between 9:00am and 11:00am, and then to gradually increase this window until it covers the entire school day.

A third tactic for increasing positive statements at times other than those selected for data collection is to use a reminder stimulus that would be present not only when recording and self-scoring are scheduled to occur, but throughout the school day. This common (to recording and non-recording environments) stimulus could be accomplished by adding public posting to the self-scoring intervention. Participating teachers could maintain their graphs on poster board, and display these charts prominently in their classrooms. The visual reminder of each day’s success and goals might serve to motivate teachers to increase their use of positive statements throughout the school day.

*Fading Experimenter Presence*

Another issue that could be addressed in future research concerns removing the experimenter from the self-scoring meetings. This would promote maintenance of the changes produced by the self-scoring procedure and would also enhance the social
validity of the results if the participating teachers were to continue using self-scoring after direct experimenter involvement was discontinued. A potential tactic for fading out the presence of the experimenter would be to have the teacher conduct self-scoring sessions daily, including graphing and posting his or her rate of positive statements, and having the experimenter conduct spot checks to insure accuracy of the teacher’s data collection. Spot checks could occur with greater frequency when first implemented, and gradually reduce to once per week.

Student Behavior Data

A third important issue for research would be to add in data on student behavior to check for a functional relationship between implementing self-scoring and improvements in student work products or on-task behavior. In addition to data collection on teacher behavior, an experimenter could collect student work products completed either across the day or during daily data collection sessions. An experimenter could also score student on-task behavior with an interval recording system, either in situ or from a videotape. Such an experiment would validate the social significance of self-scoring as a procedure to increase teachers’ positive statements.

Timing and Duration

Another modification worthy of experimentation concerns the timing and the duration of a self-scoring investigation as part of a regular (September – June) school year. Due to university and program scheduling constraints, this study began in early February and ended in the next to last week in May. All participants in this study expressed the opinion that the investigation would have been more beneficial to them had
it been implemented earlier in the school year. A future investigation could start earlier in the school year, for example in September or October, and continue to May or June. With such a schedule, an experimenter could evaluate various generalization tactics, as well as strategies to fade out his or her presence from the self-scoring process. These modifications would be necessary to assess maintenance of teachers’ increases in positive statements and decreases in negative statements over a longer period of time, and without external support.

Data Sheet Modifications

A fifth suggestion for future research concerns assessment of the effects of the data sheet that was used in this study. Because of the design and layout of this data sheet, participating teachers, through their self-scoring procedure, could build a visual display of both novel and repeated statements. This could produce a source of feedback for them that requires control in future studies and in implementation of the procedure in applied settings. A future study could assess the differences between using a form that provides specific feedback, and using a simpler form that only requires a tally or check mark for each positive statement heard. The self-scoring procedure used in this study gave detailed feedback to the participants as they filled it out. Future studies could examine the effects of using data collection systems during self-scoring that incorporate varying degrees of specificity.

Number of Self-Scoring Sessions

An additional component of this study that warrants future research is the optimum number of self-scoring meetings or sessions required to produce a lasting
change in teacher behavior. As teachers contact the naturally occurring contingencies associated with following the rules they form during self-scoring, the self-scoring meetings themselves become less important and can be reduced. Experimental manipulations could include starting with daily meetings and reducing them only when teachers meet their rate or number goals, or starting with fewer meetings and reducing them on a response-independent (without regard for the number or rate of positive statements) schedule. In the context of research on this topic, it is essential that meetings with the experimenter are eventually discontinued, in order to demonstrate that the effects can be maintained over time.

Categories and Definitions of Dependent Variables

A further suggestion for future research concerns the limitation addressed earlier regarding the categories and definitions of approval behavior that were included as dependent variables in this study. Previous studies have included non-vocal behaviors such as facial expressions, gestures, and physical contact as their dependent variables when examining the effects of teacher praise (Kazdin & Klock, 1973). A potentially beneficial replication of this study would examine the effects of self-scoring broader categories of positive teacher behaviors (including vocal and non-vocal categories) on both positive teacher behaviors and desirable student behaviors such as increased work completion and on-task behavior.

Contrived Consequences

The addition of rewards and incentives to a self-scoring intervention is a topic worthy of investigation. Rewards could include tangible or activity incentives, points, or
social recognition for participating teachers. Rewards could be delivered for meeting goals within a predetermined number of sessions, for increasing specific or novel positive statements, or for decreasing negative statements. Rewards may or may not be implemented simultaneously with self-scoring; they may be used as a secondary strategy if self-scoring alone is not effective.

Indiscriminable contingencies could be implemented that would require a certain unknown number of a particular kind of statement, or would be contingent on a particular statement (e.g., the first statement after 30 seconds) meeting a predetermined requirement (e.g., the statement would have to be novel).

An important element of introducing contrived consequences is to plan for fading those consequences over time. As teachers’ rates of positive statements increase and negative statements decrease it is important for the natural consequences (automatic reinforcement, negative reinforcement) of using a high rate of positive statements to ultimately control the behavior. In order to accomplish this, reward delivery could occur on a gradually thinning schedule, and teachers could be encouraged to keep track of their own positive statements and to reward and congratulate themselves upon meeting their goals.

Negative Statements

Another suggestion for future research is to examine the effect that self-scoring in this investigation had on participating teachers’ negative statements. Despite the fact that negative statements were not scored by the participants, negative statements decreased across participants and conditions as the study progressed. One possible explanation of
this is that the sound of each negative statement became aversive to each participant as he or she listened to negative statements during self-scoring, and thereby negative statements were punished.

A potential method for examining the effects of self-scoring positive statements on the use of negative statements would be to edit participants’ tapes to delete all negative statements before each self-scoring session (or only use 5-minute samples that do not contain any negative statements), and once data on negative statements demonstrate stability or a counter-therapeutic trend, to suspend editing (or introduce samples with negative statements) and continue to assess teachers’ rates and/or numbers of negative statements used. If data demonstrate that negative statements decrease when, and only when, teachers hear themselves making negative statements, the hypothesis that rule formation occurs when negative statements are heard would be validated. If negative statements decrease as positive statements increase, regardless of whether or not they are heard during self-scoring, then it is likely that positive statements simply produce more reinforcement for the teacher. As such, the teacher will allocate more of his or her vocal behavior to positive than to negative comments.

A variation on this study would be a reversal of the dependent variables that are scored: instead of self-scoring positive statements and setting goals to increase positive statements, participating teachers could record and set goals to reduce negative statements. A study investigating the effects of self-scoring rates of negative statements on rates of positive statements would contribute to the analysis of the effectiveness and the important components of self-scoring teaching behaviors.
Another method for assessing the effects of self-scoring positive statements on the rate of negative statements would be to include non-vocal negative statements in the dependent variables. Video recording would have to supplement every audio recording, but participants would self-score positive statements from audiotapes only. Participants would still hear negative statements, but would not see recorded negative non-vocal behaviors. The experimenter would assess the effects of self-scoring positive statements on the subsequent rate of negative statements and negative non-vocal behaviors.

Summary

This study examined the effects of teachers’ self-scoring positive statements on their subsequent rates of positive and negative statements during classroom instruction. Self-scoring consisted of listening to a 5-minute sample of one’s own instruction and recording (verbatim) the positive statements that were made and the time at which each statement was made, as measured on a digital stopwatch. After listening to each tape and recording the positive statements heard, each participating teacher graphed the rate or the number of positive statements, and set and recorded a goal for the next self-scoring session.

Four teachers employed at an urban elementary school participated in this study. For three of the four participants, results demonstrated that self-scoring produced higher rates of positive statements. All four participants emitted lower rates of negative statements after self-scoring was introduced. Of the three participants who took part in generalization recordings, three showed moderate, but short-lived increases during their generalization recording sessions, despite the fact that generalization of increased
positive statements was not specifically targeted for increase, and also despite the fact that the teachers did not know when their generalization sessions were being recorded.

The results demonstrate that although further research is needed on several components of the self-scoring intervention that was used, self-scoring is an effective and practical strategy that can be used to help teachers increase their positive interactions with their students.
LIST OF REFERENCES


APPENDIX A

INFORMATIONAL LETTER AND CONSENT FORM
February #, 2004

Dear XXXXXXXX School Teacher:

My name is Susan Silvestri and I am currently a graduate student in Special Education at The Ohio State University and a graduate assistant at the XXXXXXXX School. One of the requirements for completing my course of study is to conduct a research project. I will be conducting my research under the supervision of my faculty advisor, Dr. William L. Heward, a professor in the College of Education. I am writing to you to explain my research to you and to ask you to participate. The following is a description of the study I am planning to conduct and an explanation of your rights.

My study will examine the effects of self-scoring videotapes of classroom instruction on the frequency and type of positive statements used by teachers. Research in education has shown that positive statements (i.e., praise) from a teacher can improve student academic skills and other important classroom behaviors, and my study will look at one way to help teachers increase their positive statements. Your participation would include brief meetings and three kinds of recording during teaching: audiotaping via a wireless microphone that you would wear, videotaping via a camera on a tripod that I would operate, and videotaping via a wall-mounted wireless camera that can be activated from outside of your classroom (and therefore without your knowledge). Audiotaping will take place daily for no more than 25 minutes per session. Videotaping using both the tripod-mounted and the wall-mounted cameras will take place weekly. Tripod-mounted camera recording will occur simultaneously with one audiotaped session per week. Participation would also involve brief meetings with me in which we would discuss the recordings and count the number of positive statements made during the course of the session. The first two phases of the study are baseline and training phases in which I would audiotape and videotape during your teaching. We would not have meetings during these two phases. During the third phase meetings would occur daily at first, and would gradually decrease to once per week and would continue at once per week until the termination of the study. Meetings would take no longer than 15 minutes, and would also be audiotaped. During these meetings we would review and score short samples of your audiotapes and/or videotapes and discuss them. Sessions will be audiotaped and videotaped for the purposes of data collection and scoring only. No video or audiotapes will be shown to any other Millennium Community School participants, and no information on the video or audiotapes will be used in any kind of employee evaluation by the Millennium Community School.
If you choose to participate, every effort will be made to minimize the disruptions that audio and videotaping could potentially cause in your classroom, and to schedule recording and meeting times that are convenient for you. You are not in any way obligated to consent to participate in this research, and you will not be penalized in any way for not participating. If you do participate, you have the right to withdraw from the study at any time. During any session, if you ask to stop, the session will be terminated. Please be assured that your name will not be revealed in any publication, document, recording, computer storage or any other form of report or presentation developed from this research. Your name will not be used on any videotapes, audiotapes, or data collection forms, and all data collection forms will be coded so that only you and I will know to whom the data refer. All videotapes and audiotapes will be erased after December 31, 2004.

Attached are two copies of the consent form for Participation in Educational Research. By signing this consent form you agree to participate in this study as described in this letter. You should return a signed copy of the consent form in the stamped, return envelope and keep the second copy for your records. If you have any questions regarding this research or your rights related to participation in this research, feel free to call me at (614) 352-4028 or call Dr. Heward at (614) 292-3348. Thank you for your cooperation.

Sincerely,

Susan M. Silvestri
Ph.D. Student

William L. Heward
Professor and Faculty Advisor

Enclosures: 2 copies of Consent Form for Participation in Educational Research
Self-addressed stamped envelope
Participant Consent Form For Participation in Educational Research

I agree to participate in a research study evaluating the effects of self-scoring on teachers’ positive statements during instruction. Ms. Susan Silvestri will conduct this study under the direction of Dr. William L. Heward. The nature and purpose of this study have been explained to me, and I understand that audio and videotaping sessions will require approximately 25 minutes, five times per week. I understand that some sessions will be videotaped without my knowledge. I further understand that meetings will require approximately 15 minutes, will occur daily after baseline and training phases, and will be audiotaped. Meetings will gradually be reduced to once per week.

I also grant permission to Ms. Silvestri and Dr. Heward to use the video and audiotapes of the research sessions for data collection purposes. I understand that my identity will not be revealed to anyone not directly involved in conducting the research, or by means of publication, documentation, computer storage, or any other form of report developed from this research. Additionally I understand I may withdraw my consent for participation at any time.

__________________________      __________
Teacher’s Name                Date

__________________________      __________
Susan M. Silvestri               Date
Ph.D. Student Researcher

__________________________      __________
William L. Heward               Date
Professor and Faculty Advisor
APPENDIX B
DATA SHEETS
Participant Data Collection Form

<table>
<thead>
<tr>
<th>Participant:</th>
<th>Observer:</th>
<th>Date of session:</th>
</tr>
</thead>
</table>

Date of scoring: | Duration of observation: | Decimal: |

Transcribe positive statements, corresponding time indexes, and time indexes of repeated statements in the boxes below.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Time index</th>
<th>Record time index of repeats</th>
<th>Statement</th>
<th>Time index</th>
<th>Record time index of repeats</th>
<th>Statement</th>
<th>Time index</th>
<th>Record time index of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All positives | Total number: | Repeats: | Race: |

What did you think about the number of positive statements you recorded?

Would you do anything differently next time, or is what you’re doing working for you?

What is your goal for the next observation?
## Primary and Secondary Observer Data Collection Form

**Participant:** [Name]
**Observer:** [Name]
**Date of session:** [Date]
**Date of scoring:** [Date]
**Condition:** [Condition]
**Duration of observation:** [Duration]
**IOA recorded?** (Circle one) **Y**  **N**
**IOA data collector:** [Name]

Transcribe positive and negative statements, corresponding time indexes, and time indexes of repeated statements in the boxes below.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Generic positive (transcribe first instance)</th>
<th>Time index</th>
<th>Record time index of repeats</th>
<th>Specific positive (transcribe first instance)</th>
<th>Time index</th>
<th>Record time index of repeats</th>
<th>Negative (transcribe first instance)</th>
<th>Time index</th>
<th>Record time index of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total:**
- % repeats:
- %s repeats:
- %s repeats:

**Rate:**
- All positives: [Total number]
- % of repeats:
- Rate:
Transcribe positive statements, corresponding time indexes, and time indexes of repeated statements in the boxes below.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Time index</th>
<th>Record time index of repeats</th>
<th>Statement</th>
<th>Time index</th>
<th>Record time index of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good Job</td>
<td>0:04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THANK YOU FOR SITTING IN SILENCE,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSTING</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good Job</td>
<td>1:54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THANK YOU FOR GIVING YOUR NAME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All positives: Total number: 10
Repeats: 60%
Rate: 2

What did you think about the number of positive statements you recorded?

Would you do anything differently next time, or is what you're doing working for you?

What is your goal for the next observation? maintain - 10
Primary and Secondary Observer Data Collection Form

Participant:  
Observer:  
Date of session:  
Date of scoring:  4/23/04  
Condition:  
Duration of observation:  15:00  
Decimal:  

Transcribe positive and negative statements, corresponding time indexes, and time indexes of repeated statements in the boxes below.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Time Index</th>
<th>Record time index of repeats</th>
<th>Specific positive (transcribe first instance)</th>
<th>Time Index</th>
<th>Record time index of repeats</th>
<th>Negative (transcribe first instance)</th>
<th>Time Index</th>
<th>Record time index of repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>0.17</td>
<td>3.20</td>
<td>I like how you helped her</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beautiful</td>
<td>0.26</td>
<td>3.30</td>
<td>Thank you for not talking</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>0.56</td>
<td>4.40</td>
<td>Good - nice big word</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smart</td>
<td>0.14</td>
<td>1.10</td>
<td>You raised your hand, beautiful job</td>
<td>1.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High five</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>28</td>
<td>% repeats: 88%</td>
<td>15</td>
<td>% repeats: 63%</td>
<td>% repeats:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>1.9</td>
<td></td>
<td>2.7</td>
<td>% of repeats: 31%</td>
<td>92.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All positives</td>
<td>41</td>
<td></td>
<td>2.7</td>
<td>Rate: 2.7</td>
<td>92.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of repeats</td>
<td>31%</td>
<td></td>
<td></td>
<td>% of repeats: 31%</td>
<td>92.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

SAMPLE SELF-SCORING GRAPH
Teacher 4’s self-scoring graph
APPENDIX D

SAMPLE TRANSCRIPTS OF SELF-SCORING MEETINGS
Teacher 1

[Previous goal was 17]

Experimenter (E): (counting) 24!

Teacher 1 (T1): (clapping) Woo-Hoo!

E: Do you think you’re doing anything differently?

T1: I think I just find myself doing it more or trying to find different ways, that’s why I incorporated the sounds [letter sound review].

E: That’s good: that gave you many more opportunities. There are still so many opportunities for learning. You’ve been able to increase your positive statements very fluently, and you incorporated them into high response-rate activities that you already had.

T1: So no, I don’t think about it any more.

E: That’s good; you don’t need to think about it. You had 24 today –

T1: That is wonderful.

E: – and that should reassure you that you do not need to be thinking about it.

T1: I honestly think that if I did think about it I’d do less because I’d be so focused on it.

E: Well, whatever you’re doing, you can just keep right on doing it!

T1: So we can go to 20 [goal] for next time

E: That’s fine; I know you’d rather be successful with a lower goal than get discouraged with a big leap.

T1: Because if I don’t meet my goal I’ll be sad (laughs).
Teacher 2

[Previous goal was 7]

Experimenter (E): You had 11 today, so you met your goal. Beautiful!

Teacher 2 (T2): Good! I’m happy!

E: Good for you! Isn’t that satisfying?

T2: It’s nice to see it. I know it does a lot for me to actually see those positive statements.

E: It’s good to feedback, too. I’m glad you got to see that. What would you like to put down for your goal for next time?

T2: How about 10? I think I can do that.

E: That sounds good! I like that you’re so ambitious.
Teacher 3

[Previous goal was 4]

Experimenter (E): Today is the 29th (instructing teacher on filling out graph), and today there were 2.
Teacher 3 (T3): Okay.

E: So you can put the date, and then a 2, and then draw a line from yesterday’s session. Do you want to leave the goal where it is?
T3: Yes.

E: Like you said, you’ll just decide you’re going to do more and you will. I know that it’s easy to get distracted by some students’ behavior. You’ve named a couple of kids you know you can always count on to do what’s expected, so even if your praise is mostly to them, that’s fine. Any praise you give them might tell another student what to do.
T3: I know that praise with [student name] really helps.
Teacher 4

[First session]

Experimenter (E): Today is the 10\textsuperscript{th}, and you had 7. Awesome! So write the date here, and your 7 here. What do you want your goal to be for tomorrow?

Teacher 4 (T4): Probably…around…8.

E: Alright. That’s a good goal.

T4: (laughs) I’m sure I can meet that, it isn’t too high, but…

E: You’re right, but whatever you want to set your goal at is fine. You might want to try and find that “101 Ways to Praise a Child” list, just to give you some ideas. You can also think about your rules, and using those as very specific praise.

T4: The chart and clips help.

E: The chart is good, as long as you tell them why you’re moving their clip up. Tell them what you want them to do, and tell them when they’ve done it.
APPENDIX E

MENU OF INTERVENTION CHOICES
## Menu of Intervention Choices

<table>
<thead>
<tr>
<th>Name of Intervention</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tactile Cueing</td>
<td>This intervention requires wearing a small device called a Motivaider. It is about the size and shape of a pager, and it can be programmed to vibrate at pre-set intervals. When you feel the Motivaider® vibrate, you praise a student who is behaving appropriately.</td>
</tr>
<tr>
<td>Auditory Cueing</td>
<td>This intervention is similar to the tactile cueing intervention above, however an audible cue (e.g., a beep) will sound from a cassette player at pre-set intervals. When you hear the beep, you praise a student who is behaving appropriately.</td>
</tr>
<tr>
<td>Counter Prompting</td>
<td>With this intervention we’ll determine the number of praise statements you want to deliver during a given time period and you’ll put that number of paper clips, pennies, or other small objects into a pocket. Each time you praise a student who is behaving appropriately, you’ll move one paper clip to another pocket or another location. Your goal is to empty your pocket in the given time period.</td>
</tr>
<tr>
<td>Self-recording</td>
<td>This intervention is similar to the recording we’ve done from the videotape, except you’ll be recording each praise statement as you make it during teaching sessions. We will set up a sheet for you to record each praise statement so that you can tell how close you are to meeting your goal.</td>
</tr>
<tr>
<td>Other interventions</td>
<td>If there is a strategy that you’d like to try that is not listed here, let’s talk about it and see if we can design an intervention to best suit your needs.</td>
</tr>
</tbody>
</table>
APPENDIX F

PICTURE OF MOTIVAIDER®
APPENDIX G

TREATMENT ACCEPTABILITY RATING FORM - REVISED (MODIFIED)
Please complete the items listed below. The items should be completed by placing a check mark on the line under the question that best indicates how you feel about Ms. Silvestri’s recommendations for strategies to increase positive statements in your classroom.

1. How clear is your understanding of these strategies?
   __ __ __ __ Neutral __ __ __ Very clear
   Not at all clear

2. How acceptable do you find the strategies to be regarding your concerns about your students?
   __ __ __ __ Neutral __ __ __ Very acceptable
   Not at all acceptable

3. How willing are you to implement these strategies?
   __ __ __ __ Neutral __ __ __ Very willing
   Not at all willing

4. Given your students’ behavior issues, how reasonable do you find the strategies to be?
   __ __ __ __ Neutral __ __ __ Very reasonable
   Not at all reasonable

5. How costly will it be to implement these strategies?
   __ __ __ __ Neutral __ __ __ Very costly
   Not at all costly

6. To what extent do you think there might be disadvantages in following these strategies?
   __ __ __ __ Neutral __ __ __ Very likely
   Not at all likely

7. How likely are these strategies to make permanent improvement in your students’ behavior?
   __ __ __ __ Neutral __ __ __ Very likely
   Unlikely
8. How much time will be needed each day for you to implement these strategies?

<table>
<thead>
<tr>
<th>Little time</th>
<th>Neutral</th>
<th>Much time</th>
</tr>
</thead>
</table>

9. How confident are you that these strategies will be effective?

<table>
<thead>
<tr>
<th>Not at all confident</th>
<th>Neutral</th>
<th>Very confident</th>
</tr>
</thead>
</table>

10. Compared to students in other classrooms, how serious are your students’ behavioral difficulties?

<table>
<thead>
<tr>
<th>Not at all serious</th>
<th>Neutral</th>
<th>Very serious</th>
</tr>
</thead>
</table>

11. How disruptive will it be to your classroom to implement these strategies?

<table>
<thead>
<tr>
<th>Not at all disruptive</th>
<th>Neutral</th>
<th>Very disruptive</th>
</tr>
</thead>
</table>

12. How effective are these strategies likely to be for your students?

<table>
<thead>
<tr>
<th>Not at all effective</th>
<th>Neutral</th>
<th>Very effective</th>
</tr>
</thead>
</table>

13. How affordable are these strategies?

<table>
<thead>
<tr>
<th>Not at all affordable</th>
<th>Neutral</th>
<th>Very affordable</th>
</tr>
</thead>
</table>

14. How much do you like the proposed strategies?

<table>
<thead>
<tr>
<th>Do not like them at all</th>
<th>Neutral</th>
<th>Like them very much</th>
</tr>
</thead>
</table>

15. How willing will others in your classroom be to help implement these strategies?

<table>
<thead>
<tr>
<th>Not at all willing</th>
<th>Neutral</th>
<th>Very willing</th>
</tr>
</thead>
</table>
16. To what extent are undesirable side-effects likely to result from these strategies?

No side-effects                  Neutral            Many side-effects are likely

17. How much discomfort are your students likely to experience during the implementation of these strategies?

No discomfort at all                Neutral

Very much discomfort

18. How severe are your students’ behavioral difficulties?

Not severe at all                               Neutral

Very severe

19. How willing would you be to change your classroom routine to implement these strategies?

Not at all willing                               Neutral

Very willing

20. How well will carrying out these strategies fit into your classroom routine?

Not at all well                               Neutral

Very well
APPENDIX H

SOCIAL VALIDITY SURVEY
Social Validity Survey

Please fill in the following information:
Your job title/position (e.g., elementary school principal, 7th grade teacher, full-time undergraduate student): ______________________________________________________________

Disc Number/Name (on CD): ______________________

Video Number (circle one):  1  2  3  4

Please write a check mark in the blank that best describes your responses to the following questions.

1. How would you describe this teacher’s rapport with his or her students?
   Very   ______   ______   Neutral   ______   ______   Very
   Poor   ______   ______                      Good   ______   ______

2. How would you describe this teacher’s classroom management skills?
   Very   ______   ______   Neutral   ______   ______   Very
   Good   ______   ______                      Poor   ______   ______

3. Would you hire this teacher to work in your school?
   Absolutely   ______   ______   Neutral   ______   ______   Definitely
   Not   ______   ______                      Not   ______   ______

4. Would you want your own children to be taught by this teacher?
   Definitely   ______   ______   Neutral   ______   ______   Absolutely
   Not   ______   ______                      Not   ______   ______

5. How did watching this teacher make you feel about the instruction his or her students are receiving?
   Very   ______   ______   Neutral   ______   ______   Very
   Concerned   ______   ______                      Confident   ______   ______

Thank you!

Please write your e-mail address below if you would like more information about this study.