

THE UNIVERSITY OF MANITOBA

EFFECTS OF TOKEN REINFORCEMENT  
UPON READING COMPREHENSION AND GENERALIZATION  
WITH IMMIGRANT CHILDREN IN A NORMAL CLASSROOM

by

L. Craig Turner

A Thesis

Submitted to the Faculty of Graduate Studies  
In Partial Fulfillment of the Requirements for the Degree  
of Master of Arts

Department of Psychology

Winnipeg, Manitoba

January, 1980

EFFECTS OF TOKEN REINFORCEMENT  
UPON READING COMPREHENSION AND GENERALIZATION  
WITH IMMIGRANT CHILDREN IN A NORMAL CLASSROOM

BY

LORNE CRAIG TURNER

A thesis submitted to the Faculty of Graduate Studies of  
the University of Manitoba in partial fulfillment of the requirements  
of the degree of

MASTER OF ARTS

©1980

Permission has been granted to the LIBRARY OF THE UNIVER-  
SITY OF MANITOBA to lend or sell copies of this thesis, to  
the NATIONAL LIBRARY OF CANADA to microfilm this  
thesis and to lend or sell copies of the film, and UNIVERSITY  
MICROFILMS to publish an abstract of this thesis.

The author reserves other publication rights, and neither the  
thesis nor extensive extracts from it may be printed or other-  
wise reproduced without the author's written permission.

## ABSTRACT

Two recently immigrated children were involved in a program to increase reading comprehension. During each session the students were asked to read a specified amount of material orally and then to answer questions about the material. Following two baseline phases token reinforcement was introduced in a modified multiple baseline design across subjects and across three books. Results showed a sharp increase in the percent of correct answers to the comprehension questions and a decrease in the number of reading errors. During follow-up phases, reinforcement for each student was gradually reduced to zero for one of their three books, with no reinforcement at all being given for the other two books. The percent of correct answers to the comprehension questions remained very high across all three books; however, the number of reading errors increased slightly towards original baseline levels.

Throughout all experimental phases generalization probes were conducted across settings (i.e., from session room to home classroom) and across trainers (i.e., from experimenter to home-room teacher) for the reading comprehension behavior. Results showed that the behavior generalized well in that the generalization data were comparable to the regular session data. In addition to the number of questions answered correctly and material read correctly, several measures were taken on passages written by the students upon completion of each of the three books. No contingencies were placed on this behavior. Results showed a wide degree of fluctuation with no clear effect as a function of the experimental phases.

Prior to the commencement of the study both students scored an

average of two years below their grade level on a reading and reading comprehension test. When this test was conducted at the completion of the study, each student showed an increase of approximately four grade levels. Other results add support to the effectiveness of token reinforcement in increasing reading comprehension behavior plus presenting information on possible parameters of generalization of this behavior.

## ACKNOWLEDGEMENT

I would like to take this opportunity to thank my Advisor, Dr. Joseph J. Pear and my other committee members Dr. Garry L. Martin and Dr. Glen H. Lowther who offered me their continual support and advice over the past several years and to their constant encouragement during this study. The research study would not have been possible without the support and consent of Mr. Bob Millman, principal of Darwin School; the two home-room teachers, Mrs. Pat Wilkins and Mrs. Linda Smyrichinski who unselfishly gave me their time, and to the Resource Teacher, Mrs. Henderson, who was always around when I needed her.

Special thanks must go to Barb Roscoe whose magic typing fingers and knowledge of bureaucratic red tape made all this possible.

Finally, but most importantly, I dedicate this thesis to George and Nick. Their love of life and freedom was so contagious I never thought of this research as being work. I wish you both all the best in your new life in a new land.

TABLE OF CONTENTS

	Page
Abstract . . . . .	i
Acknowledgement . . . . .	iii
Table of Contents . . . . .	iv
List of Figures . . . . .	vi
List of Tables . . . . .	vii
Introduction . . . . .	1
Method . . . . .	6
Subjects . . . . .	6
Setting . . . . .	6
Materials . . . . .	6
General Procedures . . . . .	7
Specific Procedures . . . . .	8
Pre-baseline . . . . .	8
Baseline 1 . . . . .	8
Baseline 2 . . . . .	9
Training . . . . .	9
Follow-up . . . . .	10
Generalization probes . . . . .	10
Dependent Variables . . . . .	10
Reading comprehension test . . . . .	10
Quality of descriptive and factual information in passages . .	11
Quantity of writing . . . . .	11
Teachers' rankings of descriptive and factual information	
in the passages . . . . .	12

	Page
Teachers' subjective measure test . . . . .	12
Reliability . . . . .	12
Results . . . . .	13
Discussion . . . . .	24
References . . . . .	29
Appendix A . . . . .	35

LIST OF FIGURES

	Page
Figure 1. Mean percent correct answers to the reading comprehension questions for Student 1 . . . . .	14
Figure 2. Mean percent correct answers to the reading comprehension questions for Student 2 . . . . .	16
Figure 3. Number of reading errors for Student 1 and Student 2 across experimental phases . . . . .	17



LIST OF TABLES

	Page
Table 1. Measures of dependent variables of descriptive and factual information, and quantity of writing for passages by Student 1 and Student 2 across experimental phases . . . . .	19
Table 2. Teachers' rankings of descriptive and factual information for passages by Student 1 and Student 2 . . . . .	21
Table 3. Home-room teachers' subjective rankings of topics discussed with Student 1 and Student 2 . . . . .	22
Table 4. Pre- and post-test scores of Metropolitan Achievement Test (1971) . . . . .	23

## INTRODUCTION

The systematic manipulation of teacher attention appears to have been the first behavioral technique used in the school (Zimmerman, & Zimmerman, 1962). Teacher attention has been applied to small classes (e.g., Zimmerman & Zimmerman, 1962) and large classes (e.g., Hall, Lund & Jackson, 1968). Within the classroom setting teacher attention has been applied to one student (e.g., Kirby & Shields, 1972), as well as to the whole class (e.g., Madsen, Becker, Thomas, Koser & Plager, 1968). As well, teacher attention has also been applied to different age groups such as kindergartens (e.g., Schutte & Hopkins, 1970) and high schools (e.g., McAllister, Stachowiak, Baer & Conderman, 1969).

Research on teacher attention has usually combined the praising of appropriate behaviors and the ignoring of inappropriate behaviors. However, Madsen, Becker and Thomas (1968) studied praising and ignoring separately and found that ignoring inappropriate behaviors was not effective by itself in producing consistent acceptable behavior change. A few years later, Kirby and Shields (1972) discovered that children's attending behavior increased and decreased depending on whether the teacher's praise, contingent on appropriate behaviors, was instituted or withdrawn.

Still later, emphasis on training the teacher to train the child changed direction. Now behaviorists were more interested in directly training the student. The area of self-control or self-management became very popular.

Self-management can be divided into different categories. It usually includes one or more of the following: (1) self-determination of goals and reinforcement standards; (2) self-recording; (3) self-evaluation; and (4) self-reinforcement. Numerous programs which have used more than one of

these categories have been quite successful (e.g., Bandura & Perloff, 1967; Felixbrod, 1974; Felixbrod & O'Leary, 1973, 1974; Glynn, 1970; Lovitt & Curtiss, 1969; Turkewitz, O'Leary & Ironsmith, 1975). But not all categories by themselves have been successful. Santogrossi (1974) found self-recording to be an ineffective procedure by itself, while Santogrossi, O'Leary, Romanczyk, and Kaufman (1973) and Turkewitz et al. (1975) found self-evaluation also to be an ineffective procedure by itself in producing appropriate behavior change.

Early behavior modification programs in the school often used tangible reinforcers such as candy or money. As time went on, not only did the problems studied by behaviorists become more sophisticated and complex but so did the reinforcement systems. Along with this development behaviorists moved directly into the midst of a growing controversy over reward programs. The controversy developed not because of the effectiveness of reward programs but because of the type of reinforcement used.

Reward programs have used various reinforcers such as money, praise, feedback, tokens, etc., and generally the results showed that when reinforcement was made contingent upon specific behaviors, these behaviors increased in frequency. However, some programs which used specific types of reinforcers had conflicting results. Kuypers, Becker and O'Leary (1968) found that not all their subjects responded equally well to token reinforcement, whereas other researchers (e.g., Graubard, 1969; Harris & Sherman, 1973) found good results with this type of reinforcement system. Eventually, some researchers began to suggest that not only do some subjects not respond well to token reinforcement but they may also develop other problems such as reduced levels of interest and/or poor generalization to the natural environment (e.g., Deci,

1971, 1972; Lepper, Greene & Nesbitt, 1973). Meichenbaum, Bowers, and Ross (1968) and O'Leary, Becker, Evans, and Saudgras (1969) found that if a token reinforcement program was used during one part of the school day and not during other parts, students discriminated the intervention and non-intervention phases, which decreased the likelihood of generalization. What this meant was that the increase in academic and/or social behaviors during the token phases was not maintained during the non-token phases. Levine and Fasnacht (1974) stated that although the immediate effects of token programs may be positive, they felt that these could do more harm than good. They suggested that there should not be an attempt to use extrinsic rewards for behaviors that are of some intrinsic interest, such as classroom activities, but instead we "should first search for the natural reinforcers of problem behaviors as a significant point of intervention". Levine and Fasnacht conceded that immediate results of token systems will probably show an increase in learning but that long range results will indicate that token rewards had lead to a decrease in interest. They also felt that there was little acceptable evidence to show that generalization occurs with token programs. Finally, once token reinforcement was removed, they also felt that natural reinforcers generally did not take control over the desired behaviors.

The studies which have used token reinforcement have been numerous but they generally can be classified into three groups. The first group consists of studies which focused on one student receiving reinforcement for contingent behavior with the entire class sharing in the rewards (e.g., Brooks & Snow, 1972; Carlson, Arnold, Becker & Madsen, 1968; Coleman, 1970; Evans & Oswalt, 1968; Patterson, 1965; Rosenbaum, O'Leary & Jacob, 1975). The second group consists of studies which focused on only one student

receiving reinforcement but the behavior of the entire class was considered prior to awarding the reinforcement (e.g., Ascare & Axelrod, 1973; Barrish, Saunders & Wolf, 1969; Grandy, Madsen & DeMersseman, 1973; Harris & Sherman, 1973). The third group consists of studies which focused on the entire class receiving reinforcement (e.g., Schmidt & Ulrich, 1969; Wilson & Hopkins, 1973).

Despite the recent trend in reward programs towards the use of natural reinforcers (O'Leary & O'Leary, 1976), the concern about generalization continues. One of the most frequently used methods to facilitate generalization has been fading. This method may involve gradually decreasing the rate of reinforcement (e.g., O'Leary & Becker, 1967); gradually decreasing the number of daily reports sent home from school (e.g., Bailey, Wolf & Phillips, 1970); or gradually decreasing teacher monitoring and back-up reinforcers (e.g., Turkewitz et al., 1975). Another approach has been to program certain aspects of the treatment procedures directly into the subject's post-treatment environment (e.g., Walker & Buckley, 1972). Still another technique has been to employ feedback in the post-treatment environment in an attempt to maintain behavior (e.g., Drabman, 1973). Stokes and Baer (1977) state that a majority of behavioral researchers failed to program for generalization and instead used what the authors called a "train and hope" procedure where the behavioral researchers hoped that once appropriate behaviors had been trained, they would generalize. Stokes and Baer add that even if generalization was a possibility, a large majority of researchers failed to even determine if generalization had indeed occurred. One of their recommendations was that generalization should not be expected to occur automatically, but instead requires direct programming.

While this controversy over the use of certain reward programs continued, behaviorists also continued to expand the range of behaviors investigated. Specific behaviors studied included spelling (e.g., Rapport & Bostow, 1976), mathematics (e.g., Van Houten & Thompson, 1976), composition writing (e.g., Brigham, Graubard & Stans, 1972; Van Houten, Morrison, Jarvis & McDonald, 1974), and reading comprehension (e.g., Knapczyk & Livingston, 1974; Lahey, McNees & Brown, 1973).

Of these new areas being studied, one of the most interesting is reading comprehension. Knapczk and Livingston(1974) studied the effects of prompting question asking upon on-task and reading comprehension behaviors. The results showed that prompting was effective in initiating question asking as well as in increasing levels of reading comprehension and on-task behaviors. Lahey et al. (1973) studied reading comprehension directly with token reinforcement contingent on correct answers to comprehension questions. The authors reported a sharp increase in comprehension behavior when training was in effect. The two principal students who had previously scored at the C to F level in reading comprehension increased their comprehension skills to an A to B level. However, despite the favourable results of both these studies in increasing reading comprehension, neither attempted to take any generalization measures. Lahey et al. (1973) had no generalization data taken in the students' classrooms because all sessions were conducted in an experimental room; also there was no generalization data taken with the home-room teacher since there was only one main experimenter in the study. The authors point out in their conclusions that these types of generalization measures are very necessary and any further research in this area must consider them.

One purpose of this study was to attempt a replication of the Lahey et al.

(1973) study. Another purpose was to take generalization measures both across different settings (i.e., from session room to classroom) and across different trainers (i.e., from experimenter to home-room teacher). In addition to this research concern about token reinforcement programs and generalization there was also a practical consideration involved in the present research. The two subjects of this study were immigrant children who were having difficulty in school due to their inadequate reading comprehension skills.

#### METHOD

##### Subjects

Two brothers, 10 and 11 years old served as subjects for this study. The brothers had immigrated from Romania six months prior to the study and were enrolled in a normal school in Winnipeg. Student 1 was the 11 year old boy and he was enrolled in the fourth grade. Student 2 was the 10 year old boy and he was enrolled in the third grade. Prior to the start of the study, the students' reading and reading comprehension skills were tested on the Metropolitan Achievement Test (Durost, Bixler, Wrightstone, Prescott, & Balow, 1971). When their test scores were compared to the grade levels they were enrolled in, Student 1 was two years and three months below the expected level, while Student 2 was one year and five months below the expected level.

##### Setting

The experimental sessions were conducted in a seminar room located in the school library section. The room measured 3 m by 6 m and was equipped with a table, two chairs, and a blackboard and pegboard hanging on adjacent walls. The home-room classrooms of the two students served as the setting for the generalization probes.

##### Materials

Three books were used in this study. They were chosen from a 20-book

series from the Bodley Head Publishing Company. When the subjects were questioned on what they knew about the topics covered in these books, they stated that they were not familiar with any of the topics but they were anxious to learn. The books were The Curious World of Snakes by Alfred Leutscher (Book A); Animals of the Antarctic by F.D. Ommanney (Book B); and Animals of the Desert by J.D. Cloudsley-Thompson (Book C). The format of books were almost identical with the same number of pages, and the same number of topics throughout.

The books were located in the third and fourth grade sections of the school library. The home-room teachers and the school resource teacher felt that this was the appropriate level of difficulty to begin reading with the students.

#### General Procedures

There was a total of five experimental phases per book in this study. The phases were as follows: (1) pre-baseline; (2) baseline 1; (3) baseline 2; (4) training; and (5) follow-up. The intent of the pre-baseline phase was to take a measure of the students' knowledge about the topics to be covered in these three books prior to any actual contact. The students were therefore asked to each write one passage on each of the three books specifying what they knew about these topics. The other four experimental phases incorporated this same measuring device; however, the students had then read the books.

Individual sessions were conducted four days a week, twice a day, with each session lasting approximately 30 minutes. During each session the students were asked to read a specified amount of material in one of the books and they were then asked comprehension questions covering the material just read. They were asked to write out their answers for each question.



(See Appendix A for a list of the comprehension questions.)

Throughout all phases after pre-baseline, the percent of correct answers were recorded. During the two baseline phases correct answers were not reinforced while in the training phase correct answers did produce token reinforcement. The students worked on only one book at a time, reading them in different sequences (see next section for specific details). Each time a student finished reading a book he was asked to write a passage on that book and the other two.

The second form of daily sessional data besides the percent of correct answers to the comprehension questions was on the student's reading of the book. The book had been divided into blocks of two sentences and the experimenter recorded a correct or incorrect response for each block of sentences read.

#### Specific Procedures

The experimental procedure was a modified multiple baseline design (Baer, Wolf, & Risley, 1968) across books and across students. As was mentioned, the two students read the books in a different sequence. Student 1 read Books C, A, and B, while Student 2 read Books A, B, and C in that order. This was done to help alleviate any order effects. The experimental design incorporated five separate phases.

Pre-baseline. Prior to any contact with the books, each student was asked to write a passage on each of the three books. They were told that they had a 10-minute time limit for each passage and they were to write down everything they knew about the book's topic. Throughout this phase and remaining phases, no assistance was given to the students when writing the passages. Also, the 10-minute time limit was kept constant throughout the study.

Baseline 1. Both students read each of the three books one at a time.

A new book was not introduced until reading of the previous one had been completed. At the end of each session the student was asked the questions covering the material just read. The percent of correct answers was recorded.

The second form of data collection was on the correct or incorrect reading of the book material. Incorrect reading was defined as substituting one word for another or leaving out one or more words. The book material had been divided into blocks of two sentences and each block was recorded as being correctly or incorrectly read. Students were not penalized for incorrect pronunciation which was easily distinguishable from the substitution of a word.

Baseline 2. After all three books had been read once in Baseline 1 the student returned to his first book and the same procedures as in Baseline 1 were replicated.

Training. This phase was the first introduced immediately after completion of Baseline 2 for the student's first book. The procedures in Baselines 1 and 2 were replicated, but now the student received one token for every two blocks of two sentences mastered and one token for every comprehension question answered correctly. After completion of training on the first book, each student returned to Baseline 2 for the second book. Training was introduced for this book immediately after completion of Baseline 2. The third book followed the identical procedure after training on the second book had been completed. Again, as in previous phases, after completion of a book the student was required to write one passage on each of the three books.

During training the students were given plastic chips which could be cashed in for one of three reinforcers (i.e., free time in the gym, one-half hour drive with the experimenter, and a one-hour swim with the experimenter at a public pool). These three reinforcers were picked by the students and

ranked in terms of value. Generally each student was able to cash in on these reinforcers two to three times per training phase.

Follow-up. This phase was introduced for each student after they had completed training on all three books. Reinforcement was gradually faded out for the first book of each student. Session 1 had half the reinforcement as in training: two correct answers on the comprehension test produced one token, while every four blocks of two-sentence material read correctly produced one token. In Session 2, two correct answers on the comprehension test produced one token, while no tokens were given for correct reading of the book. In Session 3, four correct answers produced one token, while in Session 4, eight correct answers produced one token. In Session 5 there were no tokens.

For the remaining two books neither the students' correct answers to the comprehension test nor correct reading of the book produced any tokens.

Generalization probes. A total of 45 generalization probes were conducted across settings and experimenters. The generalization probe across settings involved the experimenter conducting the reading comprehension test in the classroom setting (i.e., each student's home room). The generalization probe across testers involved the student's home-room teacher conducting the comprehension test in the classroom setting.

#### Dependent Variables

There were five categories of dependent variables recorded at various intervals throughout the study. They were as follows: (1) reading comprehension; (2) quality of descriptive and factual information in the passages; (3) quantity of writing; (4) teachers' ranking of descriptive and factual information in the passages; and (5) teachers' subjective measure test.

Reading comprehension test. At the end of each session the students were

asked the questions covering the material just read and the number of correct answers was recorded. These same questions were again asked in subsequent readings of the three books. Two types of generalization probes were conducted on this test with one measuring generalization across settings from the session room to the classroom, and the other measuring generalization from the original experimenter to the student's home-room teacher.

Quality of descriptive and factual information in passages. This and the next two categories of dependent variables were used to analyze the passages written by the two students. This category consisted of three of the classification measures developed by Brigham, Graubard, and Stans (1972) which they felt measured the quality of composition writing. These measures were: (1) mechanical aspects such as length, spelling, grammar, and punctuation; (2) number of ideas; and (3) development of ideas. Each of these three measures was scored from 0 to 5 (in whole numbers) for all 78 passages written during the study. To ensure consistency between observers (i.e., observer and experimenter) sample passages were randomly chosen and used as training examples for the two observers. Rudimentary definitions were developed for these three categories and with the sample passages observers judged and discussed until there was agreement on scoring each dimension. All the passages written by the students had been typed exactly the way they were written with no identifying marks. The order in which the passages had been written was randomized to prevent the observers from knowing when they were written.

Quantity of writing. This category utilized two classification measures from Van Houten, Hill, and Parsons (1975). These composition measures were: (1) total number of words written; and (2) total number of different words. Again, the observers judged and discussed examples of these measures until

agreement was reached on scoring both measures.

Teachers' rankings of descriptive and factual information in the passages. For this category the last passage written during each of the five experimental phases was used. Four elementary school teachers (including the two home-room teachers) were given the three piles of passages corresponding to each of the three books. The passages were randomized to alleviate order effects. In this test the teachers were asked to rank the passages from 1 to 5, with 5 representing the most informative and knowledgeable passage and 1 representing the least. The two fourth grade teachers scored Student 1's passages while the two third grade teachers scored Student 2's passages. The teachers were asked to use their own means for marking papers and to rate these passages as they would rate one of their own students. The teachers worked only on the passages from one book at a time, and were blind as to the design and the intent of this study.

Teachers' subjective measure test. This final test was an attempt to have the two home-room teachers rate their student's knowledge of the three books. The test required the teacher to spend some time with her student on a random basis asking him questions about what he had been reading with the experimenter. The teacher was free to ask any questions and from the student's answers she would rank the topics discussed in the order which she felt represented the student's level of knowledge. Post-hoc comparisons were then made of the teacher's results to the experimental phases of the study.

### Reliability

For the reading comprehension data, a naive observer was given the data from three of the sessions in each of the five experimental phases. The student's answers had been typed exactly the way they were written. The observer received a copy of the questions, the answer key, the student's

answers, and the books. He was asked to grade the answers as either correct or incorrect. Interobserver reliability was calculated by dividing the number of agreements by agreements plus disagreements, and multiplying by 100.

The reliability for the first two measures on the passages (i.e., quality of descriptive and factual information, and quantity of writing) incorporated a sample of the total 78 passages written. After an observer had scored all 78 passages (once the definitions and scoring measures had been judged and discussed with the experimenter) a second observer picked 10 passages, five from each student. The only instructions this second observer received was to ensure that all three books were represented by at least one passage for each student. The experimenter was then given these passages, unaware as to which subject wrote them and as to which experimental phase they were written in. The experimenter's results were then compared to the first observer's scores on these 10 passages. Again the reliability was calculated by dividing the number of agreements by agreements plus disagreements, and multiplying by 100. Reliability measures were not taken on the other two dependent measures.

### RESULTS

Reliability measures for reading comprehension was 96% for both students (range 92% to 100%). The average of the two reliability measures on the dependent variables was 83% for Student 1 (range 76% to 100%), and 86% for Student 2 (range 73% to 100%).

Figure 1 presents the reading comprehension data for Student 1.

-----  
Insert Figure 1 about here  
-----

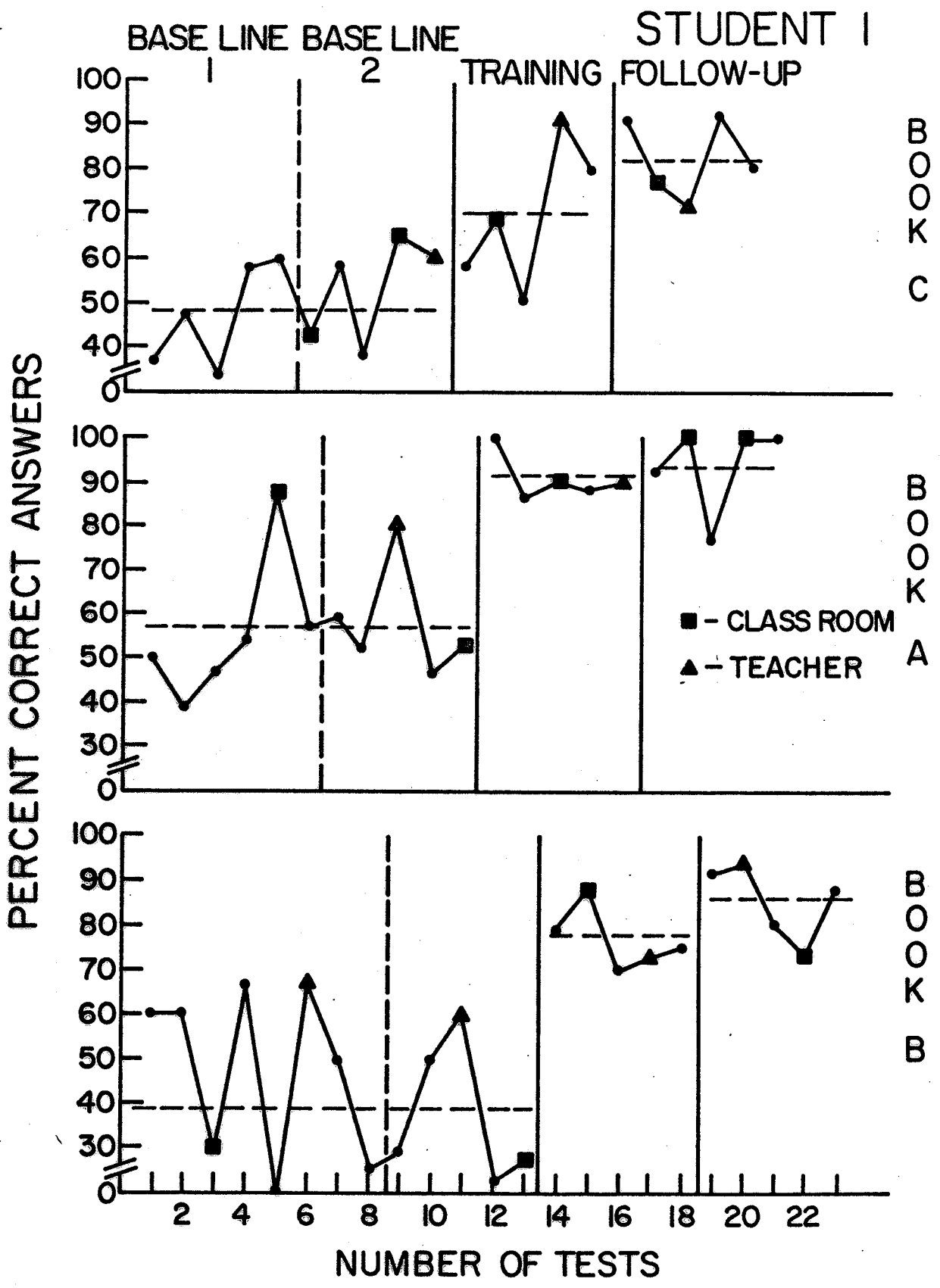


Figure 1. Mean percent correct answers to the reading comprehension questions for Student 1.