REMEDIATION OF MATHEMATICAL DEFICITS USING SELF-INSTRUCTIONAL TRAINING WITH CHECKING PROCEDURES

by

Joyce A. Pereira

A thesis submitted in partial
fulfilment of the requirements for
the degree of

MASTER OF SCIENCE IN PSYCHOLOGY

Massey University
1989

To the memory of my late father, Anthony Philip Pereira

ABSTRACT

The present study examined the relative effectiveness of three procedures for teaching long multiplication/division to seven learning-disabled adolescents: no-checking, end-checking, and multi-checking. During training, each student was taught by modelling and imitation, to verbalise self-instructions in the form of a strategy while solving the problems. The relative effects of the various checking procedures on accuracy, error rate and rate of problems completed were examined in an alternating treatments design. The best treatment was then given alone and a follow-up (a reversal) was implemented six weeks later, followed by a return to the best treatment during a final phase. Irrespective of the procedure used, the students' accuracy improved and their error rate decreased accompanied by a decline in the rate of problems completed. These effects were greatest with the multi-checking procedure for six of the seven students. Variability in performance across students indicated that the effectiveness of procedures, especially multi-checking, might be influenced by pre-skill knowledge and distractibility. Generalisation to untaught problems occurred under all procedures. Though maintenance effects were seen during the follow-up, accuracy was generally higher and more reliable with the re-implementation of the student's best checking procedure. Several hypotheses were advanced for the differential effectiveness of the procedures based on error detection and correction. Limitations of the study and some directions for further research were discussed. The findings of the study were interpreted within a radical behaviorist framework.

ACKNOWLEDGEMENTS

I would like to thank Alan Winton, who is my supervisor and friend, for all his guidance, encouragement and support in producing this thesis. His time and patience were much appreciated and his expertise was invaluable.

Special thanks are due to my mentor, Shannon Roache, for her understanding and support. My thanks also to some of the staff in the Education, Psychology and Statistics Departments for their assistance at various stages of this thesis. I would also like to thank the principal, staff, and students of Queen Elizabeth College, where this research was conducted.

Finally, my sincere thanks to my family and friends for their support and encouragement. Thanks also to Richard, who's been a pillar of strength to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Introduction

THE BEHAVIOURAL APPROACH

Despite its short history, behaviour analysis has exerted a revolutionary influence in various fields (e.g., clinical psychology, medicine, pharmacology, and law enforcement), bringing about a major reconceptualisation of psychological problems and their correction. Beginning in a relatively circumscribed area with an emphasis upon narrowly focused specific S-R relationships, the application of behaviour analysis has expanded to cover virtually the entire spectrum of human behaviour. Although the systematic use of positive and negative reinforcement in educational practice was clearly described by Jean-Marc-Gaspard Itard in 1800, the use of behaviour modification to remedy academic skill deficits is of relatively recent origin (Forness & MacMillan, 1970). Following successful interventions in the management of classroom behaviour (see K. D. O'Leary & O'Leary, 1977; Sulzer-Azaroff & Mayer, 1977), these techniques were first applied to specific academic areas in the 1960s.

Basic behavioural research has produced principles of behaviour that show how subtle changes to the environment can crucially influence the performance of each learner (J. M. Johnston & Pennypacker, 1980). Applied behaviour analysis involves "applying sometimes tentative principles of behaviour to the improvement of specific behaviours, and simultaneously evaluating whether or not any changes noted are indeed attributable to the process of application – and if so, to what parts of that process" (D. M. Baer, Wolf, & Risley, 1968, p. 91). The application of these principles to the difficulties of children requires precise evaluation of the individual's behaviour and his/her environment, precise modification of the environment, and empirical assessment of progress or change (D. M. Baer et al., 1968; Kauffman, 1975; Lovitt, 1975a, b).

The behavioural approach is appropriate for addressing a large number of educational questions, as has been documented in a number of reviews (e.g., Becker & Carnine, 1981; Carnine, 1983; Koorland, 1986). With its emphasis on precision, individualisation, and empirical evaluation, applied behaviour analysis is equally appropirate to educable mentally retarded (EMR), emotionally disturbed (ED), or learning-disabled (LD) populations (Hallahan & Kauffman, 1976).

Behaviour analysis of students' academic difficulties conceptualises them as patterns of maladaptive behaviour rather than as symptoms of some inferred processing deficit or other cognitive dysfunction. Learning difficulty is conceived as resulting from insufficient reinforcement for appropriate responses to instructional stimuli or from learning inappropriate responses to these stimuli (Dolly, 1980; Throne, 1973). Since the specific maladaptive behaviour of interest is academic, academic behaviour is the primary focus both of assessment and consequent treatment. Individual analysis and subsequent direct adjustment of antecedent and consequent events surrounding the learner's response become the primary means by which behaviourists attempt to correct learning difficulties.

Academic behaviours typically involve behaviours which have been termed 'cognitive'. Behaviour analysts using radical behaviourist philosophy (Skinner, 1974) have argued that they do consider 'cognitive' matters. However, they believe these are covert behaviours which should be dealt with in the same way as overt behaviours and should not be considered as a different kind of matter (i.e., 'mental' events). Notwithstanding this, the majority of studies in applied behaviour analysis (ABA) have tended to focus more on overt behaviour. Therefore, so called 'cognitive' areas have been dealt with more extensively by cognitive, or cognitive behaviour modification (CBM) psychologists.

At a procedural level, the experimental analysis of private events effectively involves transforming them through instructions or teaching. "What may be going on at a cognitive level can be analysed and taught at the overt level before allowing the behaviour to become covert as parts of the chain" (Deguchi, 1984, p. 88). In academic settings, one so called cognitive area where considerable behavioural research has been done is verbal behaviour (R. A. Baer, Blount, Detrich, & Stokes, 1987; Schumaker & James, 1970). Another cognitive area that has recently interested behaviour analysts, is mathematics.

The standard of education in the area of mathematics has been the source of growing concern for many. Surveys in America have indicated that the mathematical skills of pupils have been declining in recent years (Moon, 1986). In New Zealand, a survey by the International Association for Evaluation of Educational Achievement found that the mean mathematics performance at the 3rd form level was extremely low, in fact only just better than that in developing countries such as Swaziland (Forbes, 1987). Clearly there is a need, both in New Zealand and world-wide, for major improvements in the teaching of mathematics and this applies equally to normal learners as well as those that suffer from some handicap.

A sizable number of studies are concerned with the amelioration of academic skill deficits. A number of reviews on behavioural studies with students having academic difficulties are available (Gadow, Torgesen & Dahlem, 1983; Hobbs & Lahey, 1977; Kauffman, 1975; Koorland, 1986; Lloyd, 1988; Lovitt, 1975a; R. Rose, Koorland, Epstein, 1982; Treiber & Lahey, 1983). The general conclusions from these reviews are that behavioural procedures were effective. All these reviews included other academic skill areas as well as mathematics. Some did not include all the mathematics studies pertaining to the time period they were reviewing (e.g., Hobbs & Lahey, 1977; Koorland, 1986). Others have incorrectly described either the procedure of one or more mathematics studies (e.g., Gadow et al., 1983) or the information for the dependent measures and reliability (e.g., Koorland, 1986). Moreover, most of these reviews looked only at students labelled "LD" (e.g., Koorland, 1986; Lloyd, 1988). All these reviews included various academic areas with none focusing exclusively on difficulties with mathematics.

REVIEW OF LITERATURE

A comprehensive summary of behavioural studies of students with mathematical difficulties is produced in Table 1. A descriptive overview of these studies is presented below, followed by a more critical summary. This review consists of 53 studies of 1-4 experiments and covers a period of over 20 years, from 1968 to 1989. The criteria for inclusion were that the studies investigated the teaching or remediation of mathematics using ABA research methodology, although not necessarily meeting all its defining characteristics (D. M. Baer et al., 1968). Most studies involved a strict behavioural framework, although a number had a CBM orientation. In some studies two or more variations of treatments, dependent measures or reliability methods were used within the same study. If different procedures of one category were used within the same study, these were all reported separately. However, if different experiments within the same study used the same procedures for a particular category, this was counted once. This means that experiments within a study may be counted as separate studies under one category (e.g., treatment) but counted as one study under another category (e.g., academic behaviour).

Table 1

ABA research – Interventions modifying behaviour of students with mathematics/arithmetic problems

AUTHOR (DATE)	SUBJECTS	DESIGN	TREATMENTS	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Lovitt & Curtiss (1968)[9]	1M 11 yrs Behaviour disordered	A B A	Verbalisation.	Subtraction (F).	Correct rate. Error rate.	None.	Study consisted of 3 expts. which were alike in all aspects except that different problems of complexity were used.
Lovitt & Esveldt (1970) Expt. 1	1M 12 yrs Behaviour-disordered	A B A	A -Single ratio contingency (20:1). B -Multiple ratio contingency (4 different bands).	Maths (S).	Correct rate.	None.	
Expt. 2	as in expt. 1	A B A B	A -as in expt. 1 B -Multiple ratio contingency.	Addition (F).	Correct rate. Error rate.	as in expt. 1	
Expt. 3	as in expt. 1	A B A	A -as in expt. 1 B -Multiple ratio contingency (5:1), frequency of free time.	as in expt. 2	as in expt. 2	as in expt. 1	
Expt. 4	as in expt. 1	A B A B	A -as in expt. 1 B -Multiple ratio contingency (6:1).	as in expt. 2	as in expt. 2	as in expt. 1	
Kirby & Shields (1972)	1M 7th grade 13 yrs Low achiever Attentional problems	A B_1 A B_2	B_1 -Varying fixed ratio schedule of praise and immediate correctness feedback. B_2 - same as B_1 except another variation of the schedule.	Multiplication (simple & long) (F).	Correct rate. On-task behaviour.	On-task behaviour (p).	

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Conlon, Hall & Hanley (1972)	3M Low achievers 9 yrs (tutees)	A B_1 A B_2	B_1 -Peer correction procedure involving reinforcement based on time. B_2 -same as B_1 except reinforcement based on accuracy.	Arithmetic (N).	Accuracy.	None.	Measurements were taken not only of the daily academic work of the two tutees (aged 9), but also of tutor's (age not given) daily work. Parts of the procedure are unclear.
Hasazi & Hasazi (1972)	1M 8 yrs LD	A B A B	Response cost involving withdrawal of extra help (teacher attention) on numeral reversals and positive reinforcement for correct response forms.	Addition (F).	Rate of digit reversals.	Scoring (p).	Author speculated that student might be LD in discussion section.
Jenkins & Gorrafa (1972) *	12 EMR 6-11 yrs	B A B BC_1 BC_2	B -Token reinforcement. C_1 -Contingency contract. C_2 -same as C_1 except variation of contract.	Arithmetic (N).	No. of correct problems (gp. mean for each session).	Scoring (nm95).	Parts of the procedure were reported imprecisely.
McNeil, Hasazi, Muller & Knight (1972) * 2nd referral - Barbara	1F 8 yrs Low achiever	A B	Reinforcement (social).	Arithmetic (N).	Accuracy.	Scoring (nm100).	
4th referral - Kathleen	1F 8 yrs Low achiever	A B A B	Reinforcement (social).	Arithmetic (N).	Accuracy. Rate of completion of maths responses. Rate of completion of daily assigned tasks.	Dependent measures (nm100).	
Parsons (1972) 1st div.	8 students Developmentally retarded	B	Prompts, social & token reinforcement.	Pre-computational arithmetic (F).	Accuracy -overall & unprompted. Length of count.	Scoring (nm>90). Prompts, reinforcement schedule - I.V. (nm>90).	Only data for one procedure of a sequential program is presented for a student. Selected units and monitoring data on four different children are presented for the whole program. Tokens used are not mentioned.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
2nd div. (a)	as in 1st div.	B C D E	B -Finger counting procedure, prompts.	Addition & subtraction (T).	Overall accuracy. Unprompted accuracy.	as in 1st div.	Unclear about design and treatment C, D, and E.
(b)	as in 1st div.	A B A B	Strategy (circling and verbalising sign of operation) and reinforcement.	Addition & subtraction (F).	Unprompted accuracy.	as in 1st div.	Results of terminal problems shown only
3rd div.	as in 1st div.	B	Fading of prompts and increasing complexity of problems.	Story problems (F).	Unprompted accuracy.	as in 1st div.	Treatment is used across daily sessions on the story problem sequence
Pierson (1972) * Expt. 1	1M 9 yrs EMR	A B A B Follow-up.	B -Reinforcement (token).	Addition & subtraction (F).	Rate of completion. Accuracy.	Scoring (nm100).	
Pollack, Sulzer-Azaroff, Williams (1972) *	1M 16 yrs Pre-delinquent Low achiever	A B_1 A B_2 A B_1 A B_2	B_1 -Reinforcement (token) for completing weekly assignment. B_2 -Reinforcement (token) for completing daily assignment.	Arithmetic (N).	Cumulative number of assignments completed.	None.	Mean accuracy was mentioned in the description of results.
Salzberg (1972)	2 groups of 10 students. 1st group, 9-14 yrs 2nd group, 4-8 yrs. Low achievers	Multiple baseline across 2 groups of students Group 1 (B BC BCD, B BCD, BCD_2) Group 2 (B BC BCD)	Reinforcement contingencies. B -One page per day requirement minimum. C -Self-recording. D_1 -Pass one quiz per week minimum. D_2 -Pass two quizzes per week minimum.	Maths (S).	Group mean rate of quizzes passed (on each session). Rate of pages completed (voluntary homework) - for group 1 across phases.	Scoring of quizzes (p).	Before the manipulation of the reinforcement contingencies, info was collected and reported on mean rate of problems completed for gp 2 students across two phases (A B). A condition was no minimum requirement condition and B condition was one page requirement minimum.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Smith, Lovitt & Kidder (1972)[9] Expt. 1	1F 11 yrs LD	Multiple baseline across problems A B_1 B_{12} B_{123} B_{13} B_3	B_1 -Withdrawal of positive reinforcement (recess time) for C_1 errors. B_{12} -Withdrawal of positive reinforcement for C_1 and C_2 errors. B_{123} -Withdrawal of positive reinforcement for C_1, C_2, and C_3 errors. B_{13} -Withdrawal of positive reinforcement for C_1 and C_3 errors. B_3 -Withdrawal of positive reinforcement for C_3 errors.	Subtraction (F).	Accuracy.	Scoring, recording (nm100).	Error, correct, & completion rates were said to be calculated but were not mentioned.
Expt. 2	1M 10 yrs LD	A B A C A D_1 A E C A D_2	Use of teaching aids. B -Paper clips for C_{1a} problems. C -Abacus for C_{2a} problems. D_1 -Instructions for C_{1a} problems. E -Cuisinare rods for C_{3a} problems. D_2 -Instructions for C_{2a} problems.	Subtraction (F).	Accuracy.	Scoring, recording (nm100).	

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Grimm, Bijou & Parsons (1973)	2M 9 yr old Cerebral palsied. 7 yr old Hyperactive, Low achiever & having attention problems.	A_1=A_2 B A_2 A_1' A_1" C B A_2 or A_1=A_2 B A_2 A_1' A_1" B A_2	A_1 -Intermittent teacher attention (ITA), verbal praise, token reinforcement & immediate corrective feedback. A_2 -Continuous teacher attention (CTA), verbal praise, token reinforcement & immediate corrective feedback. B -Training package. A_1' -ITA, verbal praise & token reinforcement. A_1" -Extinction of all procedures. C -Verbal praise and token reinforcement for incorrect responses.	Number concept tasks (F).	Accuracy	Scoring (p).	Training package consisted of praise, token reinforcement, immediate corrective feedback, social reinforcement, verbalisation which were then gradually eliminated finally leaving the terminal reinforcers of praise and token reinforcement.
Blankenship & Lovitt (1974a)	1M 10 yrs LD	A B C Follow-up.	B -Complete a practice sheet and check answers with a calculator. C -Informing subject peer performance rates and encouraging subject to aim for the rate.	Division facts (F).	Correct rate. Error rate.	?	Secondary source.
Blankenship & Lovitt (1974b)	7 students LD	A B C BC D	B -Demonstration, cues/prompts. C -Feedback, correction of errors by student with teacher-provided prompts. D -Reinforcement.	Word problems (F).	Accuracy.	?	Secondary source. Only data of 1 subject was presented.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Fink & Carnine (1975)	4M & 6F 1st grade	A B A B	Informational feedback and graphing.	Arithmetic (?).	No. of errors.	?	Source was an abstract. Information on type of arithmetic problems and reliability info. wasn't provided.
Smith & Lovitt (1975)[9] Expt. 1	7M 8-11 yrs LD	1) A BC A^m 2) A BC A^m BCD A^m Follow-up.	Demonstration & permanent model (D & PM) B -Permanent Model. C -Demonstration. D -Instructions regarding errors.	Subtraction (F). Multiplication - simple & long (F).	Accuracy.	Timing (nm99.7). Scoring (nm>95). Recording (nm100).	Not all subjects' data was presented graphically. There were variations of experimental design within each expt. Also calculated rate data but authors stated that they did not present it.
Expt. 2	7M 8-11 yrs LD	A=D BC A^m Follow-up	D -Feedback B -as in expt. 1 C -as in expt. 1	as in expt. 1	as in expt. 1	as in expt. 1	
Expt. 3	7M 8-11 yrs LD	1) A B C A^m 2) A=D B A^m	B -as in expt. 1 C -as in expt. 1 D -as in expt. 2	as in expt. 1	as in expt. 1	as in expt. 1	Follow-up was suggested in the method but no data was presented.
Smith & Lovitt (1976) Expt. 1	3M 8-11 yrs LD	A=B C A^m	B -Reinforcement (toy) C -Demonstration	Subtraction (F). Multiplication -simple & long (F).	Accuracy.	Timings, scoring, graphing of data (nm >99).	Only data of one subject is presented graphically.
Expt. 2	7M LD	A D B A^m	D -Reinforcement (free time) B -as in expt. 1	Addition & subtraction facts (F).	Correct rate. Error rate. Median rate, median of slopes, % change of median scores for correct and incorrect problems.	as in expt. 1	Only data of one subject is presented.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Van Houten & Thompson (1976)	20 children 2nd grade Low achievers	A B A B	Explicitly timing student's maths performance for short intervals.	Addition & subtraction facts (F).	Accuracy. Overall correct rate. Local correct rate. (all 3 measures are mean performance of whole class).	Scoring (f).	
Fowler, Thomas, & Santogrossi (1977)*	1M 11 yrs	Multiple baseline across behaviours	Reinforcement (free time).	Arithmetic (N).	Accuracy.	?	Secondary source.
Smith & Fleming (1977)[g*] Expt. 1	3M & 1F LD	A B A[m]	D & PM	Addition (F). Subtraction (F). Multiplication (F).	Accuracy.	Timing, scoring, recording-graphing, data decisions (nm >89%)	Secondary source.
Expt. 2	3M LD	as in expt. 1	D & PM group instruction.	Multiplication (F).	as in expt. 1	as in expt. 1	as in expt. 1
Blankenship (1978)[g]	7M & 2F 9.0 - 11.2 yrs LD	A-B design with multiple target measures. Follow-up.	Modelling (included imitation) and feedback.	Subtraction (F).	Accuracy.	Accuracy of experimenter's reading of elapsed time (p). Accuracy (p). Teaching procedures - I.V. (p).	Only one subject's data was presented graphically. The author called it the demonstration, feedback technique. It was not stated explicitly that demonstration cards (PM) were removed after instruction.
Broughton & Lahey (1978)	33 pupils 4th & 5th grade Low achievers	Between-group design (3 contingency groups & 1 control). Within each group an A B A design was employed.	Positive reinforcement. Response cost. Mixed (both positive reinforcement & response cost).	Subtraction (F).	Accuracy. On-task behaviour (all measures were mean performances of each contingency group).	Accuracy (p). On-task behaviour (p). Teacher's behaviour -I.V. (p).	

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Hundert & Bucher (1978) Expt. 1	4M mean age = 10 yrs 4 mths Special education class	A B C B A	B -Requiring self-reporting of arithmetic problems. C -Requiring self-reporting & token reinforcement for self-reported scores on changes from prior condition.	Arithmetic (N).	Accuracy of self-scoring. Accuracy of arithmetic.	?	Source was a brief report.
Expt. 2	17 M mean age = 15 yrs 3 mths Special education class	Multiple baseline across two groups of subjects (A B C)	Reduction in checks' frequency was gradual for one group and abrupt in the other. A -same procedure as C in expt. 1 B -Maximum checks procedure (checks on accuracy of all students' self-reported scores, awarding token reinforcement for accurate self-reporting and penalties for inaccurate ones). C -Minimum checks procedure (same as B but checks done on one randomly selected student's self-reported scores each day).	as in expt. 1	as in expt. 1	as in expt. 1	as in expt. 1
Cohen, Rubin & Heinen (1979)	1M 14.2 yrs Learning & behaviour disordered	A B_1C A D B_1C B_2C C	B_1 -Token reinforcement. B_2 -Leaning of schedule of token reinforcement. C -Self-recording. D -Differential reinforcement of other behaviours.	Long division (F).	Accuracy. On-task behaviour.	Scoring (nm100). On-task behaviour (p).	Authors described design as A B A B C D

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Friedling & O'Leary (1979)* Expt. 1	7M & 1F Age range = 6yrs 10mth to 8yrs 10mth (mean age = 7yrs 7mth). Hyperactive.	Between-group design (1 experimental & 1 control). Within each group an A B_1 B_2 design was employed.	B_1 -Self-instructional training (modelling, verbalisation with fading, self-instructions, self-reinforcement). B_2 -Self-instructional training (same as B_1 plus corrective feedback, and cues).	Maths (S).	Accuracy -mean of each phase. On-task behaviour (mean of each group across each phase).	On-task behaviour (p, Kappa) Teacher attention -I.V. (p).	Control group had modelling without self-instruction. Accuracy was said to be obtained daily but was not presented graphically.
Expt. 2	same as expt. 1	Between-group design (1 experimental & 1 control). Within each group an A B design was employed.	Token reinforcement. (Teacher attention not held constant)	as in expt. 1	as in expt. 1	as in expt. 1	
Hallahan, Lloyd, Kosiewicz, Kauffman & Graves (1979) *	1M 7 yrs 11 mths Attentional problems LD	Multiple baseline across responses with reversals (A B_1 A B_1 B_2 C) Follow-up (for on-task behaviour).	B_1 -Self-monitoring with tape. B_2 -Self-monitoring without tape. C -Self-praise for being on task.	Multiplication facts (F).	Correct rate. On-task behaviour.	On-task behaviour (p).	
Varni & Henker (1979)*	3M 8-10 yrs Hyperactive, Disruptive	Multiple baseline (across behaviours)	Self-instructional (SI) training (modelling, faded verbalisation, self-instructions), Self-reinforcement (self-monitoring, graphing, self-reinforcement using tokens).	Maths (S).	Accuracy. Rate completed. On-task behaviour.	Observer drift (p). Scoring (nm100).	Secondary source. Training provided in 2 settings. After some baseline data was collected, Self-instructional training was instigated. Authors graphed data from this training in the baseline phase.
Voss (1979)	3M 8 yrs	A B Follow-up.	Drill, error correction procedure, reinforcement (praise), feedback.	Division facts (F).	Accuracy.	?	Secondary source.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Barkley, Copeland, & Sivage (1980)[*]	6M 7-10 yrs Hyperactive	A B A B	SI training (modelling, faded verbalisation, self-instructions, self-reinforcement), Self-monitoring & Self reinforcement (with 2 variable schedules of reinforcement).	Maths (N).	Rate of misbehaviour during group activity, individual work, regular school hours. On-task behaviour. Wrist & ankle actometer scores.	Rate of misbehaviour (p). On-task behaviour (p). Wrist & ankle scores (Pearson-product-moment correlation).	No measures on maths performance was collected due to problems in collection procedures.
Burgio, Whitman & Johnson (1980)[g, *]	2M & 3F 9-14 yrs Mentally retarded Attentional problems	Multiple baseline	SI training (modelling, shaping, social reinforcement, verbalisation which is not faded & within a game-like context, distraction-inoculation procedure or statements to cope with error & distractions).	Addition & subtraction (F).	SI/off-task behaviour over sessions in transfer I & II settings. Mean completion rate, mean accuracy, mean frequency of various types of SI by children across tasks.	Student behaviours (p). Event recording of specific component self-instructions (f). Rating system used in evaluating children's performance on academic tasks (p).	Maintenance data not given in results (only for 1 child). Transfer 1 setting -one-to-one situation. Transfer 2 setting - classroom situation.
Heins (1980)	4 students LD	Alternating treatments (A B) Follow-up.	B -Alternating treatments (Cued self-monitoring with recorded beeps with Noncued self-monitoring)	Maths (N).	Rate completed. On-task behaviour.	?	Source -- dissertation abstracts
Johnston, Whitman, & Johnson (1980)[g]	2M & 1F EMR (below 2 grade levels in maths)	Multiple baseline across subjects and across responses.	SI training (modelling, verbalisation with fading, shaping, self-instructions, strategy, self-controlling statements, coping statements, self-reinforcement statements).	Addition & subtraction (F).	Accuracy. Means for overall rate and accuracy for each phase.	Scoring (p).	

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Lloyd, Saltzman, & Kauffman (1981)[9*] Expt. 1	4M 8-9 yrs LD Attentional problems	Multiple baseline across subjects (B C D).	B -Pre-skills training (modelling, verbalisation, prompts with fading). C -Strategy training (modelling, use of a strategy, verbalisation with fading, redoing of whole strategy when errors are made). D -Cue training (demonstration of strategy, cues).	Multiplication facts (F).	Accuracy.	Scoring (p).	Token reinforcement during assessment periods.
Expt. 2	3M 8-9 yrs LD Attentional problems	Multiple baseline across subjects Follow-up.	Strategy training (modelling, verbalisation with fading, variable reinforcement, praise).	Multiplication & division facts (F).	Correct rate.	as in expt. 1	
Albion & Salzberg (1982)	3M & 2F 11-13 yrs Moderately mentally retarded	Multiple baseline across 4 subjects with performance of 5th subject monitored in baseline condition. Follow-up	SI training (modelling, fading of prompts, verbalisation, self-instructions, self-evaluation, self-charting, reinforcement)	Addition (F)	Correct rate. Error rate.	Scoring (p).	
Blankenship & Baumgartner (1982)[9*] Expt. 1	5M & 4F 8.1 -11.7 yrs LD	Multiple baseline across subjects { 1) A BC C CD C 2) A BC C & Follow-up }	B -Modelling (with imitation). C -Feedback. D -Token reinforcement.	Subtraction (F).	Accuracy.	Scoring, Graphing, Types of errors made, (nm100). Teaching procedures - I.V. (nm100).	C phase was used as a maintenance phase.
Expt. 2	6 from the 9 students of expt. 1 who failed criteria	A BD A Follow-up	B -Modelling (with imitation). D -Variable reinforcement (token).	Subtraction (F).	as in expt. 1	as in expt. 1	Baseline was prior maintenance phase in expt. 1. B -modelling was different from that in expt. 1. Only data for 2 subjects were presented.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Hallahan, Lloyd, Kneedler, & Marshall (1982)	1M 9 yrs LD	Alternating treatments (A B A C)	B -Alternating treatments (Self-assessed self-recording with Teacher-assessed self-recording cued by a recorded tone). C -Fading procedure where self-assessed self-recording was cued for the first part of the phase and non-cued for the second part of the phase.	Addition facts (F).	Correct rate. On-task behaviour.	Correct rate (nm100). On-task behaviour (p).	
Johnston & McLaughlin (1982)	1F 7 yrs above grade level but fail to complete assignments	Changing criterion Follow-up	Reinforcement (free time).	Computational & thought problems (T).	Accuracy. Problems completed.	Dependent measures (p). Awarding or removing free time - I.V. (p).	Length of assignments varied in terms of no. of items.
Lloyd, Hallahan, Kosiewicz, & Kneedler (1982) Expt. 1	1 M 9 years Attentional problems LD	Alternating treatments and reversal (A B A)	B -Alternating treatments (3 conditions: Self-assessment with or without self-recording, & continuing baseline)	Multiplication facts (F).	On-behaviour. Rate of movements (e.g., a numeral written).	Rate of movements (f). On-task behaviour (p).	Median scores of teacher's praise statements were given for each phase.
Expt. 2	3 students (M & F) 9-10 yrs LD	Multiple baseline and simultaneous (A B C)	B -Self-assessment. C -Self-recording.	as in expt. 1	On-task behaviour. Rate of movements.	as in expt. 1	Design: self-assessment was introduced simultaneously across subjects while self-recording was introduced in a multiple-baseline fashion.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Paine, Carnine, White & Walters (1982)[•] Expt. 1	3M 8 yrs Low achievers	For type A problems: 1) A=B=C=D E A 2) A=B=C=D E A E & Follow-up For type B problems: 1) A=B C A 2) A=B=C E A	Fading of teacher-presentation structure (use of question prompts). B -High structured board presentation. C -High structured worksheet. D -Timer contingency. E -Low structured worksheet.	Multiplication -simple & long (F).	Problem steps correct.	Scoring (p).	
Expt. 2	1F 8 yrs 3rd grade Low achiever	For both types of problems: A B A E A Follow-up.	as in expt. 1	as in expt. 1	as in expt. 1	as in expt. 1	
Schloss, Sedlak, Elliot & Smothers (1982)	1M 6th grade 11 yrs LD	Changing criterion	Token reinforcement, praise.	Maths (N).	Accuracy. Rate attempted.	Scoring for both dependent measures (f).	
Speltz, Shimamura, & McReynolds (1982) [•]	4F & 8M 7-10 yrs LD	4 reinforcement contingencies counterbalanced across four groups of three students. A_1 A_2 B C D E A_1 A_2 C B E D A_1 A_2 D E B C A_1 A_2 C D E B	B -Individualised contingency. C - All-member contingency. D -Identified responder group contingency. E -Unidentified responder group contingency.	Addition, subtraction (T).	Correct rate. Mean & standard deviation of correct rate across each phase. Social Interaction.	Social interaction (effective % agreement).	Daily data was only provided for the four lowest performing students who served as target students during one of the group contingencies and were the focus of subsequent data analysis.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Whitman & Johnson (1983)	7M & 2F mean age: 11 yrs 10 mths with range 10 yrs 2 mths to 13 yrs 8 mths EMR	Multiple baseline across 3 groups of 3 children & across responses	SI training (modelling, prompts, verbalisation with fading, shaping, self-instructions).	Addition & subtraction (F).	Accuracy. Mean rate of completion & mean accuracy for each problem type across each phase.	Scoring of dependent measures (p). Children's mastery of SI sequence (p). Children's verbalisations during assessment (p).	Generalisation was reported but data was not presented.
Rivera, Smith & Folkner (1984)[9*]	1M & 1F LD	A B A[m]	Demonstration & permanent model (D & PM).	Subtraction (F).	Accuracy.	Timing, Scoring, Recording-graphing, data decisions (nm >89%).	Secondary source.
Rooney, Edward, Polloway & Hallahan (1985)	4M age -elementary level below 2nd grade LD Attentional problems	Alternating treatments with reversals (A B A C A C A)	B -Alternating treatments (Self-monitoring of attention with Self-monitoring of academic accuracy). C -Combination of both treatments.	Maths (N)	Mean accuracy across each phase. On-task behaviour.	On-task behaviour (Harris & Lahey method).	
Van Luit & Van der Aalsvoort (1985)	2M & 2F 11yrs -13 yrs 7mth EMR	Multiple baseline (A B) Follow-up.	SI (self-instructions, problem solving strategy, modelling, faded verbalisation)	Subtraction (F).	Accuracy.	None.	Instruction using teaching aids (cuisinaire material) was given before SI.
Chiang (1986)[9]	3M & 3F 4th grade mean age of males: 10yrs 5 mths (SD=.96) mean age of females: 10 yrs 1 mth (SD= .31) LD	AE, BE_2 BE_2E_1	B -Microcomputer assisted instruction. E_1 -Assessment procedure (computer). E_2 -Assessment procedure (worksheet).	Multiplication basic facts (F).	Correct rate.	None.	Generalisation was in terms of transfer from computer to paper & pencil. A -Baseline (flashcards, instructional games).

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Montague & Bos (1986)[9]	5M & 1F age range= 15.6-18.2 LD	Multiple baseline across subjects Follow-up.	Strategy training (modelling verbalisation, paraphasing, visualising and drawing, stating the problem and hypothesising, estimating, calculating and self-checking, corrective feedback)	Verbal maths problems (F).	Accuracy. Rate of completion.	None.	
Howell, Sidorenko & Jurica (1987) Expt. 1	1M 16 yrs LD	A B A B Follow-up.	Drill & practice software.	Multiplication facts (F).	Error rate. Mean completion rate.	None.	No maintenance data was given.
Expt. 2	as in expt. 1	A B_1 A B_1B_2 Follow-up.	B_1 -Tutorial based software. B_2 -Teacher intervention.	as in expt. 1	Timed error rate. Untimed error rate.	as in expt. 1	Authors incorrectly labelled design as multiple baseline withdrawal.
Roberts, Nelson & Olson (1987)	6M & 6F 6 first graders and 6 second graders Low achievers	4 between-subject experimental conditions with 3 subjects assigned to one condition. Within one condition, a multiple-baseline across subjects was used.	Differential reinforcement (token) of use of SI, accuracy or both. SI training (modelling, strategy, prompts, reinforcement of verbalisations, self-instructions, social reinforcers leaned out). Group 1 -SI training + reinforcement for using self-instructions. Group 2 -SI training + reinforcement for accuracy. Group 3 -SI training + reinforcement for accuracy & self-instructions. Group 4 -Reinforcement for accuracy.	Addition & subtraction (F).	Accuracy. Rate of problems during which child overtly stated all self-instructions that had been taught in training.	Accuracy (nm100).	
Rivera & Smith (1988)	8 midschool students LD low achievers	Multiple baseline crossover design across 4 pairs of subjects	Demonstration-imitation-key words.	Division -long (F).	Accuracy.	Scoring (nm100). Teaching procedure -I.V. (nm100).	Reliability of teaching procedure was 89% for 1st wk which then became 100% for rest of study.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Rosenberg (1989)* Expt. 1	5M & 1F 8.6-10.1 yrs LD	Alternating treatments (adapted)	B -Alternating treatments (Structured direct-instruction with Structured direct-instruction plus homework).	Multiplication facts (F).	Correct rate of daily tasks. Rate of homework completion. Accuracy of homework assignment.	Experimenter's adherence to procedure -I.V. (nm100). Correct rate of daily tasks (nm100). Accuracy of homework assignment (nm100).	
Van Houten & Rolider (1989) Expt. 1	4 students, regular 2nd grade 6 students, regular 4th grade	Alternating treatments (A B)	B -Alternating treatments (Sequential presentation with Rapid re-presentation)	Addition facts (F). Multiplication facts (F). Division facts (F).	Accuracy.	Scoring (nm100). Experimenter's proper use of verbal praise, nonverbal approval, item presentation and use of appropriate correction procedure -I.V. (nm96).	
Expt. 2	8, regular 1st grade 2, regular 2nd grade	as in expt. 1	B -Alternating treatments (Correction with Correction plus firm reprimand)	Addition facts (F). Subtraction facts (F).	as in expt. 1	as in expt. 1	
Expt. 3	6, regular 2nd grade	as in expt. 1	B -Alternating treatments (Knee-to-knee with Desk-in-between arrangement).	Subtraction facts (F).	as in expt. 1	as in expt. 1	For both treatments all missed items were re-presented after one interspersed item and correction plus firm reprimand followed each incorrect response.

AUTHOR	SUBJECTS	DESIGN	TREATMENT	ACADEMIC BEHAVIOUR	DEPENDENT MEASURE	RELIABILITY	REMARKS
Expt. 4	2, 1st grade	as in expt. 1	B -alternating treatments (Treatment package with control).	Addition facts.	as in expt. 1	as in expt. 1	Treatment package consisted of rapid representation, correction plus reprimand procedure, and knee-to-knee teaching. Control consisted of sequential presentation, correction, and desk-in-between arrangement.

Symbols and conventions used for table 1:

Author (date)
* — studies where other academic measures as well as Maths was measured
g — studies where generalisation data was presented
g* — studies where generalisation was programmed and data presented
Expt. — experiment

Subjects
LD – learning disabled
EMR – educable mentally retarded
SD – standard deviation

Design
In describing the design, different letters are used to indicate different procedures. Furthermore, subscripts are used at times for the same letter which imply essentially the same procedure with some variation. The letter A usually denotes baseline and according to Heron, Cooper & Heward (1988) does not necessarily imply absence of instruction or treatment as such, but rather the absence of a specific independent variable of experimental interest. The equal sign (e.g., A=B) indicates functional equivalence of the two phases insofar as the dependent measures are concerned (Barlow & Hersen, 1984).
A^m denotes return to baseline to measure maintenance.

Academic behaviour
F — full description of arithmetic problems which allows for replication
S — description by source (e.g., name of text book)
T — description by type of problems (e.g., addition)
N — no description of problems

Reliability
nm — no method was mentioned
nmx(>x) — no method was mentioned but reliability was reported as x% (above x%)
p — point-by-point agreement ratio method calculated by (No. of agreements/no. of agreements + no. of disagreements) X 100
f — frequency ratio method calculated by (smaller total/larger total) X 100
I.V. — independent variable
? — reliability unknown

TREATMENTS

A variety of remediation/teaching procedures have been used. In this review these were usually divided into antecedent and consequent procedures. However, some procedures involved manipulating a mixture of antecedent and consequent variables. Consequent treatments usually included reinforcement and/or response cost.

Consequent procedures

The effects of various reinforcement strategies on academic responses were the focus of several of the earlier investigations. Reinforcement was used explicitly or implicitly in all the studies reviewed. Twelve of these used reinforcement as the principal treatment. Reinforcement was sometimes used in conjunction with another major treatment such as modelling (Blankenship & Baumgartner, 1982, expt. 2), self-instruction (R. N. Roberts, Nelson, & Olson, 1987), self-rating (Hundert & Bucher, 1978), or self-recording (Cohen, Rubin, & Heinen, 1979). Others used reinforcement either as an adjunct to other treatments or implicitly (e.g., Barkley, Copeland, & Sivage, 1980; Blankenship & Baumgartner, 1982, expt. 1; Grimm, Bijou, & Parsons, 1973; Voss, 1979, cited in Blankenship & Lilly, 1981).

Positive reinforcement was frequently used and this was generally successful in improving performance. The reinforcers used included simple verbal praise (Grimm et al., 1973; Hallahan, Lloyd, Kosiewicz, Kauffman, & Graves, 1979; McNeil, Hasazi, Muller, & Knight, 1972; Schloss, Sedlak, & Elliott, & Smothers, 1982), preselected activities during contingent free time (Broughton & Lahey, 1978; Smith & Lovitt, 1976; Speltz, Shimamura, & McReynolds, 1982) and token systems (Jenkins & Gorrafa, 1972; Lovitt & Esveldt, 1970; Parsons, 1972, 1st div., 2nd div. (b); Pierson, 1972; Pollack, Sulzer-Azaroff, & Williams, 1972; Schloss et al., 1982). Studies employing token systems usually used points, backed up by money (Cohen et al., 1979), activities/privileges (Barkley, et al., 1980; Fowler, Thomas, & Santogrossi, 1977, cited in Blankenship & Lilly, 1981; Grimm et al., 1973; Jenkins & Gorrafa, 1972; R. J. Johnston & McLaughlin, 1982; Smith & Lovitt, 1976; Speltz et al., 1982) or school supplies (Blankenship & Baumgartner, 1982; Varni & Henker, 1979).

Only a handful of studies used response-cost procedures, either alone (Smith, Lovitt & Kidder, 1972, expt. 1) or with positive reinforcement (Broughton & Lahey, 1978; Hasazi & Hasazi, 1972; Hundert & Bucher, 1978, expt. 2).

A limited number of studies examined feedback alone (Fink & Carnine, 1975) or used it as an adjunct to other treatments (Blankenship, 1978). The notion of feedback has been used in a wide variety of ways which included various antecedent and consequential events. Feedback or knowledge of results, as used in these studies, took two forms: (a) providing students with the number of correctly and incorrectly solved problems or informing them which problem was correct or incorrect (Blankenship, 1978; Fink & Carnine, 1975; Friedling & O'Leary, 1979); or (b) providing instruction or an explanation of how to solve erred problems (Blankenship & Lovitt 1974b, cited in Blankenship & Lilly, 1981; Voss, 1979, cited in Blankenship & Lilly, 1981).

The first form of feedback was usually included in most studies as a minor part of the treatment procedures (e.g., Varni & Henker, 1979). The second form of feedback, called corrective feedback (error correction) was seldom examined. It was occasionally used as a major treatment (Van Houten and Rolider, 1989, expt. 2), or as an adjunct to other procedures (e.g., Blankenship & Lovitt 1974b, cited in Blankenship & Lilly, 1981; Grimm et al., 1973; Voss, 1979, cited in Blankenship & Lilly, 1981). Techniques used in corrective feedback included modelling and/or supplying the answer (Blankenship, 1978; Grimm et al., 1973; Van Houten & Rolider, 1989, expt. 2), contingent drill (Voss, 1979, cited in Blankenship & Lilly, 1981), and the use of prompts or cues (Blankenship & Lovitt 1974b, cited in Blankenship & Lilly, 1981).

Antecedent procedures

Later ABA research has focused more on antecedent variables. More than half the studies used antecedent events as a major treatment, although often as part of a treatment package which included consequences. A number of the antecedent events were examined alone (e.g., Rivera & Smith, 1988; Rivera, Smith, & Folkner, 1984, cited in Rivera & Smith, 1987; Smith & Fleming, 1977, cited in Rivera & Smith, 1987; Smith et al., 1972, expt. 2). Antecedent variables were either modelling with or without feedback, self-monitoring, or treatment packages (e.g., self-instructional training and strategy training). A limited number of studies utilised other antecedent procedures consisting of: (a) teaching-aids (Smith et al., 1972, expt. 2); (b) computer-assisted instruction (Chiang, 1986; Howell & Sidorenko, & Jurica, 1987); (c) drill (Voss, 1979, cited in Blankenship & Lilly, 1981); (d) explicit timing of students (Van Houten & Thompson, 1976); (e) direct-instruction with or without homework assignment (Rosenberg, 1989); (f) flashcard presentation (Van Houten & Rolider, 1989, expt. 1); (g) seating arrangements (Van Houten & Rolider, 1989, expt. 3); (h) encouragement to aim for peer accuracy rates (Blankenship & Lovitt 1974a, cited in

Blankenship & Lilly, 1981); and (i) transfer of stimulus control from prompts to task-related stimuli (Parsons, 1972, 3rd div.; Paine, Carnine, White, & Walters, 1982).

However, it is sometimes difficult to categorise some studies as strictly involving antecedent variables as they contained a mixture of antecedent and consequent variables as major treatments (Blankenship & Lovitt, 1974 a, b, cited in Blankenship & Lilly, 1981; Rosenberg, 1989; Van Houten & Rolider, 1989, expt. 4) or used procedures, usually student-directed, during the emission of a response (e.g., Hallahan, et al., 1979; Heins, 1980; Lloyd, Hallahan, Kosiewicz, & Kneedler, 1982). For the purpose of this review, studies using instruction or procedures during the emission of a response or prior to assessment were categorised under antecedent variables.

Self-instructional training package

Studies using self-instructional (SI) training in the form of a treatment package emerged from 1979 onwards. Self-instruction has been defined as verbal statements to oneself which prompt, direct, or maintain behaviour (S. G. O'Leary & Dubey, 1979). It usually involved the use of self-verbalised directions to guide an individual through a series of steps that would result in the solution of a problem. The training package typically included modelling, verbalisation (which is later faded), one or more series of graduated steps for problem solution (strategies). Furthermore one or more other procedures like self-reinforcement, self-monitoring or self-evaluation, error-coping statements, and feedback were included. Eight studies used an SI training package similar to that pioneered by Meichenbaum and Goodman (1971).

There have been inconsistent findings regarding the effect of self-instructions on mathematics performance. Minimal or insignificant gains have been reported by some researchers (Burgio, Whitman, & Johnson, 1980; Friedling & O'Leary, 1979, expt. 1; Varni & Henker, 1979) while positive results have been reported by others (Albion & Salzberg, 1982; M. B. Johnston, Whitman & Johnson, 1980; R. N. Roberts et al., 1987; Van Luit & Van der Aalsvoort, 1985; Whitman & Johnston, 1983). An attempt to isolate the effective components of the treatment package was rarely made (R. N. Roberts et al., 1987).

Interventions utilising one or more components of the SI package (Grimm et al., 1973; Lovitt & Curtiss, 1968; Parsons, 1972, 2nd div. (b)) were sometimes used though no attempt was made to relate it to the SI package. Self-verbalisation was investigated in a number of studies either as a major part of the treatment or in a lesser role where the utilisation of strategies was highlighted. It usually entailed verbal statements accompanying a sequence of actions sufficient to correctly complete the problem. Sometimes verbalisation was related to a strategy (i.e., was task-specific), while at other times it involved more general statements. Results of self-verbalisation were generally positive.

Strategy training

Strategy training has been variously defined. It usually entailed teaching students some limited preskills if necessary and then teaching them how to use the preskills in sequences that lead to a solution of the problem (Lloyd, Saltzman, & Kauffman, 1981). Sometimes it consisted of a treatment package which entailed the explicit teaching of a strategy together with other components, for example, paraphrasing the problem, visualising, hypothesising, and estimating the answer (Montague & Bos, 1986).

A strategy is often used with instruction involving new skills. In general, the limited number of studies using strategy training, or procedures similar to it, have demonstrated its effectiveness. Self-verbalisation may be incorporated in the strategy which may sometimes take the form of self-guiding instructions (Grimm et al., 1973, Montague & Bos, 1986), or it may not be formally incorporated at all (Lloyd et al., 1981).

Self-monitoring

The self-monitoring literature is somewhat confusing in that there is no universally established lexicon of terms (Hallahan, Kneedler, & Lloyd, 1983). Self-monitoring, self-recording, self-assessment, and self-evaluation are some of the terms used in describing procedures by which students assess the quantity or quality of their own behaviour (S. G. O'Leary & Dubey, 1979). Self-monitoring has been defined as an individual's assessment of whether or not a target behaviour has occurred, usually followed by self-recording of the event (Nelson & Hayes, 1981). In these studies attempts were usually made to modify behaviours that could influence or be influenced by academic performance (e.g., staying on-task). Occasionally it involved evaluation of the student's progress, for example, checking whether the problem had been solved correctly (M. B. Johnston et al., 1980; Van Luit & Van der Aalsvoort, 1985).

Self-monitoring was often examined alone (Cohen, et. al., 1979; Hallahan, Lloyd, Kneedler, & Marshall, 1982; Hallahan et al., 1979; Heins, 1980; Lloyd et al., 1982; Rooney, Polloway, & Hallahan, 1985). At times it was used after training with self-instructions (Barkley et al., 1980; Varni & Henker, 1979) or as a component of an SI package (M. B. Johnston et al., 1980; Van Luit & Van der Aalsvoort, 1985). Only one study (Rooney et al., 1985) compared self-monitoring of on-task behaviour to self-monitoring of a performance variable like accuracy. While self-monitoring was successful in increasing on-task behaviours, inconsistent results were found for mathematics performance.

Demonstration/Modelling

The word 'model' has been used by most researchers to mean any antecedent stimulus that is topographically identical to the behaviour the trainer wants imitated (Cooper, Heron, & Heward, 1987). The term is often applied to "both the individual demonstrating a behaviour and the behaviour that is demonstrated" (Cooper, et al., 1987, p. 366). Sometimes it included imitation of the modelled skill by the target student(s). However, some researchers (e.g., Blankenship & Baumgartner, 1982) have used two terms for this, 'demonstration' and 'modelling', with 'modelling' meaning only 'imitation'.

Modelling was used as the principal treatment in six investigations and has been shown to be an effective technique. Sometimes, the solved problem remained as a referent while the child completed the rest of the problems independently, a procedure known as *demonstration plus permanent model* (Rivera, Smith, & Folkner, 1984, cited in Rivera & Smith, 1987; Smith & Fleming, 1977, cited in Rivera & Smith, 1987; Smith and Lovitt, 1975). This resulted in students acquiring basic computational skills which they generalised to problems for which they had received no demonstrations. In other studies an imitation component was added by having students imitate the teacher-demonstrated strategy on one problem before being allowed to compute the remaining problems independently (Blankenship, 1978; Blankenship & Baumgartner, 1982). Recently, key words were used in conjunction with modelling and this proved effective for a complex task – long division (Rivera & Smith, 1988). This procedure is quite similar to those employed in the SI studies. Modelling was also often incorporated in strategy and SI training.

PROGRAMMATIC STUDIES

There was an increasing trend for investigators to engage in programmatic research. This involved the systematic replication of successful interventions in order to define limits of applicability (R. Rose et al., 1982). One approach to such research involved the exploration of several facets of an intervention within a single study (e.g., Hallahan et al., 1982; Lovitt & Esveldt, 1970; Rosenberg, 1989). Another tactic was to explore a similar topic in several studies (e.g., Burgio et al., 1980; Friedling & O'Leary, 1979; M. B. Johnston et al., 1980).

DESIGN

An ABA (or ABAB) withdrawal design (Barlow & Hersen, 1984) or some extension of this, was mostly used during the 1970s. In some cases, a reversal of behaviour did not occur which created a problem for drawing causal inferences about the intervention (e.g., Van Houten & Thompson, 1976).

A number of designs used were not experimentally sound. These included: (a) a B design (Parsons, 1972, 1st div., 3rd div.); (b) an AB design (Blankenship, 1978; McNeil et al., 1972; 2nd referral; Voss, 1979, cited in Blankenship & Lilly, 1981); (c) other designs (e.g., A B C or A B C D) where no attempt was made to return to baseline after a treatment (e.g., Blankenship & Lovitt, 1974a, b, cited in Blankenship & Lilly, 1981); and (d) use of different evaluation procedures during baseline & intervention (Chiang, 1986). In some cases designs were incorrectly named (Howell et al., 1987) or not clearly identifiable (Parsons, 1972, 2nd div. (a)).

Recently there has been an increasing trend to use more sophisticated designs such as multiple baseline, alternating treatments or changing criterion (Barlow & Hersen, 1984). The most frequently employed design was the multiple baseline which is advantageous since irreversibility is a common problem with academic behaviour. Additional treatment phases were sometimes combined with the multiple baseline design (e.g., Blankenship & Baumgartner, 1982, expt. 2; Hallahan et al., 1979; Hundert & Bucher, 1978, expt. 2; Lloyd et al., 1981, expt. 1; Smith et al., 1972).

An alternating treatments design was employed in six studies usually to investigate self-monitoring variables (Hallahan et al., 1982; Heins, 1980; Lloyd et al., 1982, expt. 1; Rooney et al., 1985). Sometimes additional treatment phases were included (Hallahan et al., 1982; Lloyd et al., 1982, expt. 1; Rooney et al., 1985). The changing criterion design was rarely used to improve mathematics accuracy (R. J. Johnston & McLaughlin, 1982; Schloss et al., 1982).

Six investigators (Broughton & Lahey, 1978; Friedling & O'Leary, 1979; Hundert & Bucher, 1978, expt. 2; R. N. Roberts et al., 1987; Salzberg, 1972; Speltz, et al., 1982) used a between-group design, sometimes combined with a multiple-baseline design (Hundert & Bucher, 1978, expt. 2; R. N. Roberts et al., 1987; Salzberg, 1972). However, in only one case were both individual-subject data and group analysis of variance presented (R. N. Roberts et al., 1987).

MAINTENANCE AND GENERALISATION

Generally little attention was paid to maintenance (generalisation across time) or generalisation across problems and settings in the 1970s. In the 1980s, while all three types of generalisation (across problems, settings and time) were investigated only two received more than cursory attention.

Follow-up probes were frequently used (18 studies) in assessing maintenance. Others used a return to baseline conditions to test for maintenance (Rivera et al., 1984, cited in Rivera & Smith, 1987; Smith & Fleming, 1977, cited in Rivera & Smith, 1987; Smith & Lovitt, 1975, 1976). In these studies the maintenance period ranged from a few days to four months with only seven groups of investigators (Hallahan et al., 1982; Heins, 1980; Montague & Bos, 1986; Paine et al., 1982; Pierson, 1972; Smith & Lovitt, 1975, expt. 1; Van Luit & Van der Aalsvoort, 1985) measuring it for more than six weeks.

Much attention has also been focused on increasing students' generalisation across problems. This is commendable since the question of generalisation has become central in considering the efficacy of behavioural techniques (Stokes & Baer, 1977). Thirteen studies probed for generalisation of treatment effects, yet only four (Blankenship & Baumgartner, 1982, expt. 2; Burgio et al., 1980; Lloyd et al., 1981; Rivera et al., 1984, cited in Rivera & Smith, 1987) explicitly programmed for generalisation. Various methods were used to program generalisation across problems. These included the use of a verbal cue and variable reinforcement (Blankenship & Baumgartner, 1982, expt. 2), targeting the most difficult problem within a problem cluster for instruction (Rivera et al., 1984, cited in Rivera & Smith, 1987), multiple exemplars (Burgio et al., 1980) or explicit teaching of strategies for applying preskills (Lloyd et al., 1981). Generalisation across settings was rarely investigated (Burgio et al., 1980).

ACADEMIC BEHAVIOUR

Although mathematics (maths) behaviours were evaluated in nearly all the studies, other topics were frequently of primary interest to the investigator. These included: (a) other behaviours such as on-task time (e.g., Cohen et al., 1979); (b) generalisation across problems (e.g., Rivera et al., 1984, cited in Rivera & Smith, 1987); and (c) relationship among academic areas or subskills (e.g., Lloyd et al., 1981).

In describing mathematical behaviours a full description of the mathematical problems was usually included. However, ten studies did not provide any description of problems used other than "arithmetic" or "maths" and six others provided descriptions inadequate for replication. Investigators reporting the use of basic facts were categorised as giving a full description. Arithmetic operations, especially addition and subtraction, being frequently examined, enjoyed the widest range of interventions with multiplication next. Division was the least examined arithmetic behaviour in terms of studies and intervention techniques.

Fewer studies investigated other mathematical areas, for example, word problems (Blankenship & Lovitt, 1974b, cited in Blankenship and Lilly, 1981; Montague & Bos, 1986; Paine et al., 1982; Parsons, 1972, 3rd div.). Fractions and algebra were areas that were not addressed by any of the reviewed studies. As most of the subjects had mathematical deficits the level of difficulty of most of the maths tasks was slightly below that expected for 'normal' age-peers.

DEPENDENT MEASURES

Most researchers included data on some if not all individual subjects. Others (e.g., Broughton & Lahey, 1978; Van Houten & Thompson, 1976) examined the performance of individual subjects but presented the subject-data in a single figure.

All the reviewed studies evaluated maths performance except for two (Barkley et al., 1980; Heins, 1980). The measures of maths performance commonly used were accuracy (usually percentage correct) or time-based measures such as frequency (rate) of correct and incorrect responses. Though frequency measures are more sensitive and precise than accuracy measures (White & Haring, 1976), 26 investigators used accuracy measures only to measure maths performance.

In measuring academic (maths) performance, several studies included only one rate measure (Chiang, 1986; Hallahan et al., 1982; Hallahan et al., 1979; Heins, 1980; Howell et al., 1987, expt. 2; Kirby & Shields, 1972; Lloyd et al., 1981; expt. 2; Lovitt & Esveldt, 1970, expt. 1; Speltz et al., 1982). This made it extremely difficult to make any judgements about the subjects' accuracy. Only 13 studies reported either an accuracy and a rate measure, or any two rate measures which is required for adequate assessment of treatment effects (e.g., Albion & Salzberg, 1982; Howell et al., 1987, expt. 1; Varni & Henker, 1979).

The use of other types of measures to measure maths performance included rate of digit reversals (Hasazi & Hasazi, 1972), rate of quizzes passed (Salzberg, 1972), unprompted accuracy (Parsons, 1972), percentage of problem steps correct (Paine et al., 1982) or rate of movements in which a numeral was written (Lloyd et al., 1982, expt. 1).

Along with the usual measures of maths performance, a number of studies included measures for: (a) other types of academic subjects (Barkley et al., 1980; Burgio et al., 1980; Fowler et al., 1977, cited in Blankenship & Lilly, 1981; Friedling & O'Leary, 1979; Hallahan et al., 1979; McNeil et al., 1972; Pierson, 1972; Pollack et al., 1972; Rosenberg, 1989; Speltz et al., 1982; Varni & Henker, 1979); (b) attending behaviour (Broughton & Lahey, 1978; Burgio et al., 1980; Cohen et al., 1979; Friedling & O'Leary, 1979; Hallahan et al., 1982; Hallahan et al., 1979; Heins, 1980; Kirby & Shields, 1972; Lloyd et al., 1982; Rooney et al., 1985; Varni & Henker, 1979); (c) social interaction (Speltz et al., 1982); (d) misbehaviours (Barkley et al., 1980); or (e) SI behaviour (Burgio et al., 1980; R. N. Roberts et al., 1987).

RELIABILITY

Reliability was not reported consistently across studies. About one-third of the investigators did not test for reliability of maths performance, although reliability measures for non-academic behaviours were included for some (Barkley et al., 1980; Friedling & O'Leary, 1979; Hallahan et al., 1979; Kirby and Shields, 1972; Rooney et al., 1985; Speltz et al., 1982). Of those who tested for reliability, several provided no data about reliability procedures for maths performance measures, although they reported reliability values of 100% or less (e.g., Blankenship & Baumgartner, 1982; Cohen et al., 1979; Jenkins & Gorrafa, 1972; R. N. Roberts et al., 1987; Smith & Lovitt, 1975, 1976). However, some of these studies included reliability procedures and values for non academic measures (Cohen et al., 1979; Hallahan et al., 1983; Varni & Henker, 1979).

A number of investigators used two or more reliability methods for different measures within the same study. Reliability procedures and values for at least one academic or nonacademic dependent measure were reported in 23 studies, most of which were conducted in the 1980s. Two types of calculations were used to determine the reliability of observers. The majority (20 studies) employed the point-by-point agreement ratio method (though sometimes not for all dependent measures used within the study). Other methods used included: (a) the frequency method (Burgio et al., 1980; Lloyd et al., 1982; Schloss et al., 1982; Van Houten & Thompson, 1976); (b) the Harris & Lahey method (Rooney et al., 1985); (c) effective percentage agreement method (Speltz et al., 1982); or (d) the Pearson product-moment correlation (Barkley et al., 1980). Kappa was rarely used (Friedling & O'Leary, 1979, expt. 1).

The observed reliability values, when provided were generally within the acceptable range, usually being greater than 90%. They often reached 100% for permanent-product measures but were somewhat lower (73%-95%) for event or interval-recording of on-task behaviours.

Even though the effective control of the independent variable is an integral requirement of proper experimentation, adequate attempts to ensure this were made in only nine studies (Blankenship, 1978; Blankenship & Baumgartner, 1982; Broughton & Lahey, 1978; Friedling & O'Leary, 1979; R. J. Johnston & McLaughlin, 1982; Parsons et al., 1972; Rivera & Smith, 1988; Rosenberg, 1989; Van Houten & Rolider, 1989).

SUBJECTS

Most investigators reported the subjects' gender. Male subjects outnumbered female subjects approximately 3:1. Some 60% of the studies used students aged 6-10 years whilst only 8% (Howell et al., 1987; Hundert & Bucher, 1978, expt. 2; Montague & Bos, 1986; Pollack et al., 1972) used students aged 15 years and above.

The students were usually either LD, mentally retarded or low achievers. LD students were usually used and a number of them were reported as having attentional problems. The next largest category of students were low-achievers. These included those who were either in the same grade level or below the grade level of their peers, and had poor performance and difficulty in maths. Some of these may have been categorised as LD if tests had been done. Subjects used in other studies were categorised as behaviour disordered, developmentally retarded, cerebral-palsied, hyperactive, or normal.

CONCLUSION

There has been progress in several aspects of behavioural research on remediation of mathematical skill deficiencies. Designs in recent studies have become more appropriate and sophisticated. More investigators are using innovations such as more than one type of design, or combining designs to test the effects of one intervention. There is also an increase in the amount of programmatic research and the pursuit of maintenance and generalisation issues. Furthermore, subjects of both sexes are being included in current investigations. Although a variety of reinforcement procedures were sometimes used, recent studies have focused more on antecedent events.

Remediation techniques have become more varied with an increasing trend to use student-directed procedures. These student-directed procedures generally fell under the rubric of CBM and were often vague about the behavioural principles involved. Moreover, there was a neglect of techniques other than modelling, self-instruction and reinforcement. Too many treatment techniques were 'packaged' together without any attempt of component analysis (e.g., in self-instruction).

While there is an increasing trend for studies to examine generalisation across problems it is still largely "train and hope" (Stokes & Baer, 1977, p. 350) and thus there exists a continuing need for studies to program generalisation (Bursuck & Epstein, 1987). Furthermore, generalisation to other settings has been rarely examined.

Treatment effects were usually inadequately assessed through the use of only one dependent measure. There was also a failure to validate hierarchies of component skills (Gadow et al., 1983). Only one study examined the relationship of a subskill on a complex task like long multiplication/division.

Procedural explanations have improved, but more must be done with regard to subject descriptions. Even though a large number of investigations used LD, diagnostic criteria were often not specified and data pertaining to the discrepancy between ability and achievement were sometimes omitted. Moreover, the issue of a widely acceptable definition of learning disabilities remains unresolved with the literature characterised by disagreements over etiology and remediation. The majority of the studies were only concerned with elementary-aged children and 'normal' subjects were rarely used.

Inadequate descriptions of the mathematical problems occurred in some studies. There was usually a focus on arithmetic operations especially basic facts, addition, and subtraction, with relatively few or no applications to other mathematical areas such as word problems, and fractions.

Though more recently investigators are reporting reliability, more control for observer reactivity and careful documentation of reliability procedures are needed. Furthermore there is also a need for more adequate attempts to ensure the effective control of the independent variable. The two reliability methods most commonly employed are point-by-point agreement ratio and frequency, but these have the problem of high expected/chance levels of agreement (Hartmann, 1977) and may produce inflated or ambiguous estimates (Kazdin, 1982). Though several methods have been proposed to handle this problem (e.g., Kappa), they have yet to be widely adopted.

THE PRESENT STUDY

This review has shown that a number of techniques like modelling, and strategy training have been effective in the remediation or teaching of maths. However, there is a considerable need for studies to examine the effectiveness of a number of other remediation or teaching techniques used for mathematics. The present study was an attempt to look at one of these techniques. The following sections review more closely aspects of treatment, generalisation, design, subjects, and academic behaviour which are directly relevant to the present study.

TREATMENTS

Error-correction

One such technique is corrective feedback or error correction. It is well established that responses are maximally affected by consequential stimuli, occurring with minimum delays. Hence when using a contingent stimulus to alter a response, it is typically presented as soon as possible after the response. However, with complex behaviour sequences, providing an immediate consequence for one response may affect the emission of other important responses. Such 'feedback' (as it is frequently termed) is seen to comprise both antecedent and consequential events. Feedback alone may not be effective in improving accuracy when students are acquiring a skill (Blankenship & Lovitt, 1974c, cited in Blankenship & Lilly, 1981). Feedback, when combined with instruction, appears to be more effective when teaching new skills, or after a student has acquired a skill and is striving to become proficient in it (Blankenship & Lilly, 1981).

Some type of corrective feedback (error correction) is critical in any instructional situation and is used often in the classroom. The effectiveness of a variety of error-correction procedures is well documented in the reading literature (Jenkins & Larson, 1979; T. L. Rose, McEntire, & Dowdy, 1982; Singh, Winton, & Singh, 1985). A positive relationship has also been demonstrated between the correction of students' errors and increased maths performance (Carnine, 1977, cited in Rosenberg, 1986; Van Houten & Rolider, 1989). For example, Carnine (1977, cited in Rosenberg, 1986) taught preschool children several sets of arithmetic facts, first without corrections and then with corrections. Accuracy on the dependent measures averaged 55% higher during the correction phases than during the no-correction phases. Following error correction with a decelerating consequence such as a mild reprimand may further decrease the likelihood of errors. For instance, Van Houten & Rolider (1989) demonstrated that error correction with a reprimand yielded higher accuracy than correction with no reprimand.

Techniques used by the instructor in correcting errors included: (a) supplying the answer (Carnine, 1977, cited in Rosenberg, 1986; Van Houten & Rolider, 1989); and (b) contingent drill on errors (Voss, 1979, cited in Blankenship & Lilly, 1981). Another approach was to require the student to correct their errors with the help of teacher-provided prompts (Blankenship & Lovitt, 1974b, cited in Blankenship & Lilly, 1981).

Although the efficacy of error correction as a teaching or remediation technique has been known for some time, observations of maths instruction in the classroom and an analysis of current mathematics texts indicate that a consistent error correction procedure is not common. Generally corrective feedback is external, that is, provided by the teacher. However, corrective feedback can be internal (Grimes, 1981) as when the student detects his/her error without any external prompting. One way of doing this is by the student checking. Such student-directed procedures are required especially with the emphasis by some behaviour therapists (e.g., Kazdin, 1975) for switching the locus of control over an individual's behaviour from external agents to the individual himself. These student-directed procedures would also save time for the teacher. The present study examined student-directed error-correction by using two types of checking procedures and a no-checking procedure. It is reasonable to expect that students would correct errors detected with checking.

Self-instructional training

Effectiveness of SI training

To investigate error-correction, it has to be incorporated in the context of an effective instructional technique. Furthermore the experimental analysis of checking, which usually involves private events, needs to be transformed effectively at a procedural level as instructions, and investigated at the overt level. For this, SI training seems appropriate since in teaching students one or more strategies, and in requiring the students to verbalise what they are doing, it allows the effects of that teaching to be observed.

A number of recent maths remediation studies have used SI training similar to that used by Meichenbaum & Goodman (1971). Though its effectiveness has been documented in several studies (Albion & Salzberg, 1982; M. B. Johnston et al., 1980; Van Luit & Van der Aalsvoort, 1985; Whitman & Johnston, 1983), minimal or insignificant gains were reported in a number (Burgio et al., 1980; Friedling & O'Leary, 1979; Varni & Henker, 1979). The inconsistency of results across studies may be attributed, at least partly, to the nature of the self-instructions used (Albion & Salzberg, 1982). Most of the studies used SI routines that differed to some extent from those used by Meichenbaum & Goodman (1971). Instead of providing a detailed explanation of the training procedure, some reports (e.g., Leon & Pepe, 1983; Varni & Henker, 1979) merely indicated that they followed Meichenbaum & Goodman's (1971) training procedures. This was usually not sufficient to determine if both general and specific problem solving strategies were included, if both accuracy and strategy use were monitored explicitly and if coping with failure and self-reinforcement statements were included. Furthermore few attempts have been made to isolate the effective components of SI training packages with most studies employing group designs (e.g., Barling, 1980; Leon & Pepe, 1983). Only one study (R. N. Roberts et al., 1987) using ABA analysis was found.

Components of self-instruction

Essentially the three major components of self-instruction are modelling, verbalisation and one or more strategies. In addition one or more self-statements involving problem definition, focusing attention, planning and guiding responses, self-reinforcement, self-monitoring (self-evaluation), coping and error correction are used. However, it is sometimes difficult to strictly categorise studies utilising one or two components of the SI routine; for instance when self-verbalisation was investigated as a major part of treatment (Lovitt & Curtiss, 1968) or when self-verbalisation was of lesser importance but the utilisation of strategies was highlighted (Lloyd et al., 1981).

Modelling appeared to have been a vital component for training students to use self-guiding instructions in all the SI studies reviewed. It is "certain that modelling is a useful technique" in the systematic instruction of students in most academic skills especially as a single teaching procedure (Lloyd, 1980, p. 61). However, its relative effectiveness as a component of the SI training "package" has yet to be investigated. Modelling is efficacious when accuracy is in question and though it can be applied with groups of students, its effects seem to be more positive and consistent when applied individually (Rivera & Smith, 1987). Moreover, comparable results have been found across entire instructional sequences for addition, subtraction, and multiplication. However, it might be nonfacilitating for a task involving long division (Rivera, et al., 1984, cited in Rivera & Smith, 1987) due to the complexity of such tasks (Rivera & Smith, 1988).

Self-verbalisation has usually entailed either task-specific or general verbal statements, accompanying a sequence of actions sufficient to correctly complete the problem. Its utility in training maths skills has been demonstrated in some cases (e.g., Grimm et al., 1973; Lovitt & Curtiss, 1968; Parsons, 1972), but not in others (e.g., Cullinan, Epstein & Lloyd, 1978). R. N. Roberts et al. (1987) suggested that overt verbalisation may not be a necessary component of the training regimen. However, in order for a person other than the speaker to analyse SI behaviour, overt verbalisation is required, in order to make the procedure observable.

What appears to be most important to the success of SI training is that it teaches students specific strategies for working on academic problems. This was demonstrated by R. N. Roberts et al. (1987) who used a full SI regimen, and then differentially reinforced the use of self-instruction only, accuracy only, or both self-instruction and accuracy. The self-instruction was found to be no more effective than the strategy training plus reinforcement of arithmetic accuracy. Thus, they concluded the effectiveness of self-instruction seems to lie in it providing step-by-step, clear instructions (a strategy), rather than any contribution to assumed cognitive processes. In other words the strategy provides important discriminative stimuli for the emission of appropriate written responses.

In general, the few studies using strategy training (e.g., Lloyd et al., 1981; Montague & Bos, 1986) or procedures similar to it (e.g., Grimm et al., 1973; Smith & Lovitt, 1975) have demonstrated its effectiveness. However, Gerber (1987) has suggested that the concept of strategy appears to drift in response to the conceptual position of the writer, being sometimes described as an overt behaviour, sometimes as self-verbalisation, and sometimes mental activity presumed to relate to information handling and transformation.

Results from self-monitoring research indicate that increased time on-task may produce slight but inconsistent improvement (Hallahan et al., 1979), or no improvement at all in maths performance (Lloyd et al., 1982). Snider (1987) concluded that effects are generally not as consistent or dramatic for increasing academic responding as for improving on-task behaviour. Graden, Thurlow, & Ysseldyke (1983) have argued that it is not simply attending but making an active academic response that is crucial to learning. Thus self-monitoring of academic performance variables such as accuracy appears to be a promising research direction. However, such studies are practically non-existent, and only one study (Rooney et al., 1985) was found that compared self-monitoring of on-task behaviour to self-monitoring of accuracy. The self-monitoring component could be expected to be valuable for students who have consistently experienced failure in maths and who may therefore have become "learned-helpless" (Gresham, 1987). This is because in monitoring oneself, the relationship between actions and outcomes are highlighted (Ryan, Weed, & Short, 1986). In the present study, the SI components of modelling in conjunction with a verbalised strategy were used to teach LD students a complex task – either long multiplication or long division. The strategy which involved a series of graduated steps to guide students in solving the problem comprised the instructional techniques of demonstration, imitation, and verbalised self-instructions.

The present study was also designed to contribute to the limited body of research in this area by investigating checking (error-monitoring) procedures, which is one form of self-monitoring or self-evaluation of the performance variable. It investigated the relative efficacy of two types of checking procedures, end-checking (ec), and multi-checking (mc) and a no-checking (nc) procedure, applied to students solving a complex computational task. To date no research has investigated self-evaluation along these lines. The current investigation was therefore conceived as an initial exploration of student-directed (rather than teacher-directed) error-correction procedures using verbalised self-instructions.

Self-instruction and behaviour analysis

While these SI procedures have usually been successful, the results have often been analysed in cognitive terms and the underlying behavioural principles remain vague.

SI programs have been labelled 'cognitive' by some researchers (Brigham, 1980; S. G. O'Leary & Dubey, 1979) as verbal behaviour is included in the procedure. According to Skinner (1957), in considering the properties of verbal behaviour, there are at least two systems of responses, one based upon the other, with the upper level only being understood in terms of its relation to the lower. The notion of an inner self is an attempt to illustrate the fact that when behaviour is compounded in this way, the upper system seems to guide or change the lower. But the controlling system is itself also behaviour.

In some of the self-instruction studies reviewed above, self-instructions were considered to be manipulations of functional cognitive events. This view has two fundamental flaws in the eyes of radical behaviourists (Deguchi, 1984). First, data in these studies often only show a close *correlation* between certain verbal instructions and subsequent overt behaviour and do not justify positing a causal relationship between cognitive operations and behaviour. Moreover, what were manipulated at the procedural level were not cognitive events but types of instructions. Second, hypothesising cognitive processes has limited utility, and tends to stop focus on the ultimate causes of behaviour – environmental events outside the organism which are directly accessible and effectively manipulable (Skinner, 1974, p. 10).

According to Deguchi,

> A behavior analysis of the self-instructional procedure concedes that self-instructive behavior can be made part of a behavioral chain controlling other behavior through direct teaching of self-instructive behavior, that is, as a function of certain prior environmental histories. In this view, private events are not inferred substitutes for the reinforcement history because they are demonstrably products of it and demonstrably functional. Those events are parts of the chain that follows the immediate environmental antecedents whose functions have been created by their historical interactions with reinforcement contingencies (1984, p. 87).

Theoretical models

The relationship between speech and other behaviour proposed by Skinner (1963), Vygotsky (1934), and Luria (1961), has received empirical support from recent research on human operant performance (Lowe, 1979; Lowe, Harzem & Bagshaw, 1978; Lowe, Harzem, & Hughes, 1978).

Other than the model of verbal control of motor behaviour described by Luria (1961), there are few theoretical models describing the mechanisms underlying the reported effectiveness of SI training (R. N. Roberts & Nelson, 1983). One hypothesis is that the instructional component of self-verbalisation may act as a discriminative stimulus directly increasing the probability of a correct response (S. G. O'Leary & Dubey, 1979). In order to do this, it is necessary to reinforce a correspondence between specific self-instructions and nonverbal behaviour rather than simply reinforcing the student's self-instructions. Evidence that such a procedure is effective comes from recent behavioural research on verbal-nonverbal correspondence (e.g., Israel & O'Leary, 1973; Karoly & Dirks, 1977; Risley, 1977; Rogers-Warren & Baer, 1976).

Another hypothesis is that self-instruction indirectly affects performance by focusing attention on the task (S. G. O'Leary & Dubey, 1979). Several studies providing tentative support for this hypothesis (Hartig & Kanfer, 1973; Meichenbaum & Goodman, 1969; Palkes, Stewart, & Freedman, 1972) have shown that overt (audible) self-instructions are more effective than covert (silent) self-instructions. However, for the "attention" function to be clearly established, it is necessary to demonstrate that overt verbalisation improves attention to the task as well as better task performance.

Thus a cognitive interpretation of self-instruction is neither essential nor desirable. A behaviour analysis is both possible and useful for delineating important processes. In line with this view the present study regarded the use of self-instruction to deficit academic skills as a process that involved chains of behaviours which could be strengthened through reinforcement. Accurate performance involved emitting chains of appropriate verbal and motor (written) behaviours with important discriminative stimuli being provided by self-instruction. The three procedures investigated in the present study differed in the way the links of the chain were checked for errors. For instance in nc, no checking of errors occurred in any link while under ec, checking for errors occurred only when the whole chain was completed. Under mc, errors were usually checked after emission of each link. It would be reasonable to expect that mc would result in higher accuracy than either nc or ec as checking each link in the chain would maximise the probability of detecting errors, allowing for their correction. It would also be expected that ec would result in higher accuracy than nc as some form of checking is better than none.

GENERALISATION

SI training has little functional value if there is no generalisation of any behaviour gains (Cole & Kazdin, 1980). Generalisation of SI training has been shown to be limited (Schweitzer & Sulzer-Azaroff, 1988). When generalisation was not observed, it is frequently unclear whether it was because the student's verbalisation did not control the appropriate responding or because the student failed to produce the self-instructions. When generalised effects did occur, it is often unclear to what degree covert self-instructions were responsible for these positive changes. With a few notable exceptions (e.g., Burgio et al., 1980), researchers have seldom attempted to verify that verbalisation has actually acquired the desired control of the relevant overt behaviours as a result of training, or inspected systematically the student's use of self-instructions in the generalisation environment. To investigate this, the present study, unlike other studies, did not fade verbalisation, but continued them throughout.

The literature also shows that while a number of studies did examine generalisation, they did not program it. A few studies (Rivera et al., 1984, cited in Rivera & Smith, 1987; Smith, 1978, cited in Rivera & Smith, 1987; Smith & Fleming, 1977, cited in Rivera & Smith, 1987) have shown that targeting the most difficult problem type within a cluster may enhance generalisation to both similar and dissimilar types of problems. In one such study, Smith (1978, cited in Rivera & Smith, 1987) examined generalisation across problems and instructional sequencing of two groups of LD students. Task-analysed and sequenced groupings of arithmetic problem types were categorised. One group received instruction on the most difficult problem type within the sequenced problems and was tested on the easiest problems (i.e., a difficult-to-easy sequence) whilst a second group was instructed on the easiest problems and was tested on the most difficult problems (i.e., an easy-to-difficult sequence). For both groups the application of D&PM led to mastery of taught problems, but generalisation results differed. Generalisation to uninstructed problem types occurred more often for students given the difficult-to-easy sequence than for those given the easy-to-difficult sequence. Further, when the sequence of instruction was reversed, those students who did not generalise in the easy-to-difficult sequence did so in the difficult-to-easy sequence.

The present study investigated such generalisation by targeting a problem of medium difficulty and testing with problems of equal, less and greater difficulty.

DESIGN

More sophisticated designs have been increasingly used in the maths literature. One of these is the alternating treatments design (Barlow & Hersen, 1984), which is appropriate for studying academic behaviours which, once learned, are irreversible. It allows the relative strengths of several treatment alternatives to be determined by rapidly alternating between treatments. This design has several advantages. The rapid alternation of treatments minimises time-correlated artifacts that might occur if each intervention was tested serially as in reversal or multiple baseline designs. Stimulus conditions other than the programmed treatments (e.g., time of day) which might influence the data are counterbalanced so that their effects can be separated from those of the programmed intervention. Furthermore this design requires less time to conduct. The present study used an alternating treatments design in investigating the three procedures.

Unfortunately the alternating treatments design is susceptible to multiple treatment interference (Barlow & Hayes, 1979; Barlow & Hersen, 1984). However, by following the alternating treatments phase with a phase in which only one treatment is administered, the experimenter can assess the effects of that treatment in isolation (Cooper et al., 1987).

SUBJECTS

The present lack of consensus over the definition of learning disabilities and the procedures for operationalising the LD definitions, makes the identification of an LD sample difficult. This problem is compounded in New Zealand by the absence of a formally identified LD group (Chapman, 1985) since the state provides no official remedial teaching services.

Notwithstanding these difficulties, an attempt was made to identify a sample of LD students that had characteristics which at least commensurate with those frequently reported in the literature (e.g., Kavale & Nye, 1981; Shepard, Smith & Vojir, 1983). The procedures adopted for identifying LD children in this study were similar to those used by Chapman & Associates (1984, 1985, 1988) in their studies with LD students in New Zealand.

ACADEMIC BEHAVIOUR

The teaching of computational arithmetic comprises a major portion of the maths curriculum for LD students (Cawley, Fitzmaurice, Shaw, Kahn, & Bates, 1978; Cox, 1975a) because many of them demonstrate substantial computational deficits (Fleischner, Garnett, & Shepard, 1982; Kauffman, 1981; Mercer, 1979) across their school years (McLeod & Armstrong, 1982). Thus although it is generally accepted that arithmetic should be mastered, instructors must identify approaches that expedite such mastery to allow time for other components of maths instruction. Keeping in mind the mathematical "interest level" of beginning secondary students, the present study used a complex task -long multiplication or division.

In summary the present study utilised an alternating treatments design to study two types of checking procedures and a no-checking procedure within the context of an SI package. The present study also attempted to apply a behaviour analysis to the findings.

The major aims of the present study were:

1. to determine the relative efficacy of the three procedures, using verbalised self-instructions, on the performance of third-form students with mathematical deficits involving a complex computational problem, long multiplication or long division.

2. to examine the effect of training problems of intermediate complexity on generalisation to problems of similar, less and greater complexity.

Method

SUBJECTS

Subjects came from regular form 3 classes of a Palmerston North public secondary school. Seven students aged 13-14 years (five females and two males) served as subjects. Three were Maoris and the rest were Caucasian. All came from low-middle income groups. Hearing, vision (with correction) and emotional development were reported as normal.

SUBJECT SELECTION

Thirty four potential subjects were referred by their classroom teachers as having academic problems in mathematics. These students were then screened by teachers using the Abbreviated Teacher Rating Scale (ATRS), a global behaviour rating scale. Children scoring in the impulsive range, that is, above a mean ATRS rating of 1.5 per item, were eliminated from the study. The remaining were then given further tests by the experimenter to identify those considered to be maths learning disabled.

The basis for final selection were that they met the following criteria:

1. a maths Progressive Achievement Test (PAT) score at or below the 17th percentile,

2. scores greater than or equal to the 20th percentile on at least two of the other PAT subtests,

3. a Wechsler Intelligence Scale for Children - Revised (WISC-R) full scale IQ of 90 or above,

4. had the necessary preskills (see pre-baseline assessment),

5. scored less than 20% accuracy in long multiplication or long division test problems

Table 2 summarises each student's age, gender, ethnicity, attention/impulsivity, achievement, and IQ data. Fictitious names were used to conceal the identity of the subjects.

PREBASELINE-ASSESSMENT

On the basis of preskills assessment, five students (3 female and 2 male) were selected for long multiplication training and two students (both female) were selected for long division training.

Assessment was a two-step process. The students were first assessed to determine their abilities with basic computational facts and skills involving addition, subtraction, multiplication, and division. Students' performance was assessed with worksheets containing problems representative of the entire sequence of each computational process.

The second step involved student assessment on task-analysed sequences of skills in long multiplication and division. There was a wide range of computational abilities among the students. It should be noted that all the students have had varying degrees of multiplication and division instruction in their previous school years depending on their ability level. Students' comments suggested that instruction was based mainly on demonstration and practice.

Students doing long multiplication required the following preskills:

1. an understanding and mastery of at least 85% accuracy for addition,

2. an understanding and mastery of at least 70% accuracy for basic multiplication facts,

3. mastery of the basic ideas of place values through the 100000s.

Students doing long division required the following preskills:

1. an understanding and mastery of at least 85% accuracy for addition and subtraction,

2. an understanding and mastery of at least 70% accuracy for basic multiplication and division facts,

3. facility in multiplying and dividing by powers of 10,

4. facility in rounding,

5. facility with estimating such as 6 x __ = 8248 (either 1000 or 10000),

6. mastery of the basic ideas of place values through the 100000s.

Table 3 summarises each student's performance levels of basic computational facts and skills for the four basic computational operations.

The subjects were treated in accordance with the ethical standards of the American Psychological Association (1983).

SETTING

The research was conducted in a vacant cubicle in the school's remedial reading room.

MATERIALS

ASSESSMENT INSTRUMENTS

The identification of LD students varies with different researchers. However, most LD definitions include the notion of discrepancy between ability and achievement. Indeed, many suggest that the discrepancy factor is the common denominator of learning disabilities (e.g., Mercer, 1983). Thus the two key components usually used in operationalising the discrepancy factor, are estimates of intelligence (to establish the presence of at least normal intellectual functioning) and measures of achievement (to establish specific areas of learning difficulty) (Chapman, 1985). Both these components were used in this study.

Table 2

Description of subjects: summary of gender, age, ethnicity, Conner's rating for impulsivity, achievement (vocabulary, comprehension, listening, mathematics) and IQ (verbal, performance, and full) scores for each student

					ACHIEVEMENT SCORES[b]				INTELLIGENCE SCORES[c]		
SUBJECT	GENDER	AGE	ETHNICITY[a]	CONNER'S RATING	VOCAB	COMPRE	LISTN	MATHS	VOCAB IQ	PERF IQ	FULL SCORE
SUE	F	13.3	C	1.3	18	24	54	6	88	91	91
JOE	M	13.9	M	0.8	3	21	30	13	85	101	91
VERA	F	13.6	C	0.2	84	51	60	7	90	104	96
KEITH	M	13.5	M	0.2	36	37	61	10	101	104	102
MARY	F	14.1	C	0.4	36	20	14	15	88	93	90
JANE	F	13.7	C	0.4	40	28	54	16	92	100	95
JOAN	F	13.6	M	0.1	20	11	36	17	87	98	91

[a]C = Caucasian; M = Maori. [b]As measured by the PAT. [c]As measured by the WISC-R.

Table 3

Accuracy on basic facts (BF) and computational skills (CS) for each student in each of the four basic mathematical operations

SUBJECT	ACCURACY (%) FOR EACH TYPE OF OPERATION							
	ADDITION		SUBTRACTION		MULTIPLICATION		DIVISION	
	BF	CS	BF	CS	BF	CS	BF	CS
SUE	100.0	100.0	100.0	58.4	70.7	7.0	62.0	10.5
JOE	100.0	100.0	100.0	65.0	73.3	10.0	70.0	12.0
VERA	100.0	100.0	100.0	73.0	70.0	10.0	68.0	11.0
KEITH	100.0	100.0	100.0	76.0	86.0	13.0	83.0	15.0
MARY	100.0	100.0	100.0	72.0	76.2	13.6	74.0	17.0
JANE	100.0	100.0	100.0	85.2	97.4	86.0	90.0	20.0
JOAN	100.0	100.0	100.0	85.8	96.0	90.0	94.8	23.5

The WISC-R (Wechsler, 1974) was used as the measure of general intellectual functioning. It has been described as the best standardised test and the most widely used test for school-age students (Taylor, 1984). It correlates highly with other measures of intelligence (Kaufman, 1976; Taylor, 1984). The subscales are moderately reliable and highly intercorrelated and the derived verbal and performance scales are even more reliable and highly intercorrelated (Mitchell, 1985). The WISC-R has also been shown to predict academic achievement equally well for blacks and whites and to be relatively free from item bias (Reynolds, 1982).

Achievement was measured by four of the PAT Series, namely Reading Comprehension (Elley & Reid, 1969), Reading Vocabulary (Elley & Reid, 1969), Listening Comprehension (Elley & Reid, 1971) & Mathematics (Reid & Hughes, 1974). These tests were developed specifically for use in New Zealand and thus the test content is largely in line with the curricula in most New Zealand (NZ) schools. They are probably the four most frequently used standardised measures of achievement in the NZ school system (Chapman, St. George, & Van Kraayenoord, 1984). Their split-half reliability coefficients are above 0.85 and the tests are described as having medium to high validity in NZ (Elley & Reid, 1969, 1971; Reid & Hughes, 1974).

The ATRS (Conners, 1973) is a 10-item, 4 point Likert scale (0-3), derived from the full-scale Conners Teacher Rating Scale, CTRS (Conners, 1969). It categorises the degree to which children are perceived by their teachers as impulsive and disruptive. On the basis of a normative study children scoring above a mean of 1.5 per item are said to be impulsive (Sprague, Cohen, & Werry, 1974). The CTRS, test-retest reliability scores ranged from 0.70 to 0.90 (Conners, 1973). Satisfactory correlations have been reported between the ATRS and both the hyperactivity factor and the mean of all factors of the CTRS (Sprague et al., 1974).

WORKSHEETS

Random numbers were used to generate 25 multiplication/division worksheets although some selection of the numbers was made to meet specifications for the worksheet. Each worksheet had two sets of twelve problems. Each set had four problems with single-digit multipliers (sdm) or single-digit divisors (sdd), four with double-digit multipliers (ddm) or double-digit divisors (ddd), and four with triple-digit multipliers (tdm) or triple-digit divisors (tdd). Within each set, the order of presentation was randomised with the stipulation that no more than two problems having the same number of digits in the multiplier/divisor occurred successively and that at least one of each problem type occurred every seven problems. This was to ensure that students completed a comparable number of problems in each cluster. All worksheets had comparable levels of difficulty. A different worksheet was used in each session. A sample multiplication/division worksheet is provided in Appendix 1.

Multiplication worksheets

The four sdm problems had: (a) one with a three-digit multiplicand; (b) two with four- and/or five-digit multiplicands with no zeros; and (c) one with a four- or five-digit multiplicand containing one or two zeros in the second and/or third and/or fourth and/or fifth place of the multiplicand.

The four ddm problems had: (a) one with a two-digit multiplicand with no zeros; (b) one with a three-digit multiplicand with or without a medial zero; (c) one with two zeros in the second and either third or fourth place of the three/four-digit multiplicand; and (d) one with a four-digit multiplicand with no zeros.

The four tdm problems had: (a) one with a three-digit multiplicand with or without a medial zero; (b) one with a four-digit multiplicand with no zeros; and (c) two with a four-digit multiplicand which contained a zero in the second and/or third place.

Division worksheets

The four sdd problems had: (a) one with a three-digit dividend; (b) one with a four-digit dividend; (c) one with a five-digit dividend; and (d) one with a three- or four- or five-digit dividend which had a zero in the second or third place of the dividend. Of these four problems, two problems yielded at least one zero in the quotient.

The ddd problems had: (a) one with a three-digit dividend; (b) one with a four-digit dividend; and (c) two with a five-digit dividend, where at least one contained a zero in the dividend. Of these four problems, one or two problems yielded at least one zero in the quotient.

The tdm problems had: (a) one with a four-digit dividend; and (b) three with a five-digit dividend where at least one contained a zero in the dividend. Of these four problems, one or two problems yielded at least one zero in the quotient.

Worksheets used during the alternating treatments phase

To allow for better comparison between the three procedures in the alternating treatments (AT) phase, three variations of a multiplication worksheet were used within each treatment block. The numerals of each problem in one worksheet was contained in a problem in the other two worksheets but with their place values varied, for example, 4635 x 23; 6543 x 32; 5436 x 23. The three division worksheets within each treatment block were made comparable so that each variation of one problem gave roughly similar quotients, for example, $324\overline{)6156} = 19$; $325\overline{)6175} = 19$ and $323\overline{)5814} = 18$.

DEMONSTRATION PROBLEMS

The solving of problems was demonstrated during the AT phase as part of the instructional procedure. None of these problems appeared identically on the students' worksheets.

For multiplication, three different problems were used, one for each of the first three sessions of the AT phase. Each of the problems had a four-digit multiplicand with a zero in the tens place and a two-digit multiplier.

For division, three different problems were used, one for each of the first three sessions of the AT phase. Each of the problems had a five-digit dividend and a two-digit divisor which did not yield any zeros in the quotient. For the first session of the third block of the AT phase, one problem with a five-digit dividend and a two-digit divisor which yielded a zero in the tens place of the quotient was used.

FOOD REINFORCERS

Mars bars were provided for attendance and punctuality on five random occasions for each student throughout the study.

DEPENDENT MEASURES

Various academic performance measures of accuracy and rate were calculated from each 20-minute assessment period. Accuracy measures were expressed as a percentage and rate measures as a frequency for the assessment period (i.e., number of problems per 20 minutes).

1. Overall accuracy was defined as the number of problems correctly solved divided by the total number of problems completed, multiplied by 100.

2. Accuracy for each of the three problem clusters (single-, double-and triple- digit multiplier/divisor problems) was defined as the number of problems in that problem cluster correctly solved divided by the number of problems completed which belonged to that particular problem cluster, multiplied by 100.

3. Overall completion rate was defined as the total number of problems completed.

4. Completion rate for each problem cluster was defined as the number of problems completed of that particular problem cluster.

5. Overall error rate was defined as the total number of problems incorrectly solved.

6. Error rate for each problem cluster was defined as the total number of problems incorrectly solved of that particular problem cluster.

DESIGN

For each student, sessions usually occurred twice daily (separated by at least one hour), every week day except when holidays or special school events occurred. The times of sessions varied each day because of the constraints of the school timetable. During all phases, sessions involved a 20-minute assessment period. Additionally, in some sessions of the AT phase, training in the two SI checking procedures (using ddm/ddd problems) was given for 25-40 minutes prior to the assessment period. All training sessions were tape-recorded for the full length of the session. All assessment sessions were tape-recorded for 12-15 minutes of the 20-minute period, with the part of the session recorded varying randomly across sessions. Graphical feedback was provided in each assessment period. This was done by the experimenter marking the worksheet in front of the student immediately after assessment, and graphing the score to show progress. Criteria for phase changes depended on students' performance with ddm/ddd problems.

An alternating treatments design was utilised to compare the effects of a no-checking procedure and two checking procedures on the accuracy and rate of multiplication/division problems. Originally the study was to include the following phases: baseline, verbalisation, alternating treatments (Barlow & Hersen, 1984), best treatment, and a follow-up phase after

6 weeks to test for maintenance. However, as a result of the treatment procedure chosen by the student during the follow-up, the study was extended to include another phase, a second best treatment phase which was similar to the first best treatment phase.

BASELINE (BL)

During this phase, the experimenter handed out the assessment sheets with the following instructions: "These papers have some very difficult problems. Many of them will be hard for you to do. Don't worry, just try your best to complete correctly as many as you can." Students were also told to attempt the problems in the order presented.

The students completed the worksheets quietly and no assistance in solving the problems was provided. Baseline conditions were scheduled for four sessions provided that stability of ddm/ddd problems (as denoted by no wide fluctuations in data) was achieved.

VERBALISATION (VB)

This phase, was similar to the BL phase except that the students were instructed to verbalise what they were doing as they solved the problem. They were also told that the experimenter was interested in knowing how they solved the problem and that verbalisation also improved concentration. Verbalisation conditions were scheduled for four sessions provided that stability of ddm/ddd problems was achieved.

ALTERNATING TREATMENTS (AT)

Following completion of the VB phase an AT phase was instigated. In each of the three sessions of the first block, prior to the assessment period, training in the two SI checking procedures was given for 25-40 minutes. Throughout this phase, one of three treatment procedures was instigated during each assessment period. The order of presentation of the treatment procedures was randomised initially and then counterbalanced, with the provision that each procedure occurred once in every block of three sessions. Counterbalancing was employed to the treatment procedures to minimise carryover effects and control for order effects (Barlow & Hersen, 1984). The procedures were (a) no-checking (nc); (b) end-checking (ec); and (c) multi-checking (mc).

To increase the discriminability of the three treatment procedures (Kazdin, 1982), each session was prefaced with specific instructions concerning the procedure to be implemented. Furthermore the worksheets were given different letters (A, B, C) to indicate the different procedures to be used.

The nc treatment procedure consisted of students verbalising a strategy using self-instructions, to solve the problem. The ec treatment procedure was similar to the nc procedure, but in addition required checking of most links after the problem was solved. The mc treatment procedure was also similar to the nc procedure except that, in addition, students were required to recheck their response to every link while solving the problem, usually by using the inverse operation for those doing division or the same operation in the reverse direction for those doing multiplication. Checking normally occurred after each computational link but on occasions had to be after two or three links as in the subtraction links for division students. This was because it was natural to complete a set of links before checking. To have multichecking immediately after each link would have artificially broken the natural continuity of the computation, and would have made a relatively short process inordinately long. Details of the mc and ec procedures for multiplication and division are provided in Appendix 2.

Verbalisation was required by students throughout all training and assessment periods. However, following adverse comments from the students concerning verbalisation, the assessment procedure was changed from the second to last block of this phase. Students were required to verbalise the procedure only during the first twelve minutes of the assessment session but then point to the numerals in the problem without verbalising during the last eight minutes. Pointing to the numerals was another way to ensure that the students were following the procedure. The students took roughly the same amount of time pointing as they did for verbalising.

This phase was scheduled for at least 6 blocks, with the proviso that one treatment emerged as the most effective. Accuracy on ddm/ddd problems was used to determine the most effective of the three treatments.

BEST TREATMENT 1 (BT1)

In the following phase, the most effective of the three treatment procedures was implemented for three sessions.

FOLLOW-UP (FU)

Six weeks after the last session of the BT1 phase, an FU phase was implemented for three sessions to test for maintenance. Conditions were quite similar to those used during VB except that students were told to complete the worksheets using any of the three procedures they had learnt during the AT phase. Moreover, unlike the VB phase, verbalisation was required for the first 12 min., but were required to point to the numerals in the problem during the last 8 min., while solving the problem. This was to enable the experimenter to determine which checking procedure they were using.

BEST TREATMENT 2 (BT2)

This phase was identical to BT1 and also lasted for 3 sessions.

PROCEDURES

Self-instruction consisted of a series of steps (a strategy) verbalised by the subject to solve the long multiplication/division problems. The specific instructions employed were formulated after a careful task analysis of the maths algorithms involved in long multiplication/division. To ensure that the language used in the self-instruction was familiar to the students, teachers were interviewed regarding terminology used currently in the classroom and pilot testing of the instructions was done with some students, prior to the study.

Students were trained only on ddm/ddd problems. Generalisation effects were examined for non-instructed problems (sdm/sdd and tdm/tdd). In addition to the training given to all students in the first block, division students were given a further training session on the first session of the third block of the AT phase, to introduce another problem of greater difficulty (one producing a zero in the quotient). Modelling of the two SI checking procedures (i.e., ec and mc) was provided during each training session irrespective of whether it was an nc, ec, or mc assessment session. A full training script for one multiplication and one division training session is provided in Appendix 2.

During each training session, only one multiplication/division problem was used in the demonstration-imitation procedure for both ec and mc, though the problems differed across the three sessions. The same problem was used in each training session so as not to confound the different treatment procedures with the introduction of a new problem.
As in the BL phase, a rationale for verbalising was again provided. Before modelling the treatment procedure, the experimenter said "This is the procedure where we end-

check/multi-check (check as we go)". The experimenter demonstrated one checking procedure (randomly assigned across subjects), and then demonstrated the other checking procedure, verbalising the problem and performing the corresponding motor behaviour as the student observed. Modelling of the experimenter detecting and correcting an error was included. After each of the demonstrations, the student performed the same task independently verbalising the instructions and performing the corresponding steps of the algorithm with prompting by the experimenter when required. When prompting, the experimenter did not supply answers or correct any computational errors made by the student. She only ensured that the student followed the correct procedure. If the student had difficulty, did not respond for four seconds, or requested help with the demonstrated steps of the strategy, the experimenter asked the following questions: (a) What is the problem? (b) What have you just done? and (c) What should you do next? The experimenter did not provide answers to the questions nor was the strategy demonstrated again for that session. The answers to most of these questions required students to think about the self-instructions used at each step of the strategy, in order to continue computing the problem. If the child made an error in the self-instructions, the experimenter said "Stop, What did you just do?", and "What should you do next?" Again the experimenter never supplied any answer.

The student's verbalisation was reinforced with verbal praise if it corresponded with the ongoing performance of the algorithm. The student was allowed to verbalise the self-instructions in his/her own words as long as these approximated those taught by the experimenter and contained those elements necessary for correct completion of the problem. Reinforcement was also given if students corrected their detected errors while checking.

Prompts were also given by the experimenter throughout the whole assessment period, for all sessions except those in the BL, VB, and FU phases, when required. This was to maintain use of self-instructions not only for the problems belonging to the instructed class (ddm/ddd) but for all those problems belonging to the uninstructed classes for which students spontaneously generalised the procedure.

RELIABILITY AND VALIDITY

The experimenter and a graduate student independently assessed the permanent-product data (worksheets) from all assessment sessions for each student. The interobserver reliability (IOR) was calculated for all dependent measures separately. Rater agreements on both accuracy and rate were calculated on a problem-by-problem basis. Reliability was assessed using the point-by-point agreement ratio method (Kazdin, 1982), by dividing the number of agreements by the number of agreements plus disagreements, multiplied by 100. Kappa (Barlow & Hersen, 1984) was also calculated again for both the accuracy and the rate. In all instances, reliability coefficients were 100% and Kappa equalled 1.

The experimenter and a graduate student (same one as used for IOR assessment) independently assessed tapes (randomly) from all training and assessment sessions for each student. Two types of procedural validity were assessed. First, whether the experimenter had correctly implemented each procedure during training was assessed. The graduate student was familiarised with the procedure with the help of a sheet outlining the experimental procedures, initially. In all instances and for all students, both assessors found the procedures were correctly implemented.

Second, the appropriateness of the experimenter's instructional prompts (verbalisations) during the training and assessment periods was evaluated. Experimenter verbalisations were categorised into (a) correct prompts; (b) incorrect prompts which were either supply prompts (where students' errors were corrected or their questions answered) or no prompts (where prompts were not given when students used the incorrect procedure); and (c) other verbalisations (verbalisations not related to instruction).

Each experimental verbalisation was categorised and recorded using 15-s intervals in a partial-interval recording (Cooper et al., 1987) for five minutes of each audiotaped training and assessment period. If more than one prompt occurred in any interval both were recorded as occurring. Rater agreement was calculated on an interval-by-interval basis using point-by-point agreement ratio method, and Kappa was also determined.

For the training sessions, all experimenter prompts were correct prompts for all students. For assessment sessions, on average, 88.7% of experimenter verbalisations were correct prompts, 0.3% were supply prompts, 6.3% were no prompts and 4.6% were other verbalisations, with the multiplication students. With the division students, on average, 80% were correct prompts, 10% were supply prompts, 0% no prompts and 10% other verbalisations. Thus validity was high. There was complete interobserver agreement on the occurrence and categorisation of experimenter verbalisations for all training and assessment sessions, which resulted in a reliability value of 100% and a Kappa value of 1.

Students' mastery of self-instructions was assessed during the third training session for five minutes on each procedure. An experimental guideline sheet was used which listed the steps of the self-instructions taught to the students for correct solving of problems. The student's mastery of these self-instructions was determined by the number of steps actually verbalised divided by the number of steps needed to correctly solve the problems, multiplied by 100. Those steps which required teacher prompts were considered as not being verbalised by the student. Mastery of self-instructions for multiplication students ranged from 85% to 94% for mc, and 92% to 96% for ec. For division students, mastery of self-instructions was 80% to 86% for mc and 86% to 92% for ec. In all instances, point-by-point agreement ratio reliability was 100% for all students.

Results

ACADEMIC PERFORMANCE

Two main measures of each student's academic performance are presented for each 20 minute assessment period throughout all phases of the experiment. The measures included accuracy (%) and rate of problem completion (number of problems/20 min.) for each of the three problem clusters as well as for total problems. Figures 1 - 7 show students' accuracy scores for problems with double-digit multipliers/divisors (ddm/ddd) in the upper left, single-digit multipliers/divisors (sdm/sdd) in the upper right, triple-digit multipliers/divisors (tdm/tdd) in the lower left and for all problems completed (overall) in the lower right. Figures 8-14 show the completion rates for each student, again for ddm/ddd, sdm/sdd, and tdm/tdd problems and for all problems separately. In the AT phase of figures 1 - 14, a block of three sessions, one for each treatment procedure is shown above each point on the abscissa. Tables 4 - 10 present the mean accuracy scores, mean completion rates and mean error rates of each student in each phase. Tables 12 - 18 (which are contained in Appendix 3) present the raw data of each student for each assessment period throughout all phases of the experiment.

In the following sections, an analysis of the data from each phase is presented, first for the accuracy scores and then the completion rates. Where there were patterns shown by most students, these are outlined in an introductory paragraph. Then the results for the various problem clusters are presented, beginning with the cluster from the class of problems used for instruction during the AT phase (ddm/ddd), then the uninstructed clusters (sdm/sdd and tdm/tdd) and finally the combined results from all the clusters (overall). A brief description of mean error rates is also included.

BASELINE (BL)

During baseline, accuracy on sdm/sdd problems varied but was 0% on multidigit multipler/divisor problems (ddm/ddd and tdm/tdd) for all students but one.

On ddm/ddd problems, all students scored 0% for all four sessions except Mary, who had a mean accuracy of 12.5%. All students commenced at levels of 20% to 40% on sdm/sdd problems. On sdm problems, accuracy scores increased across the phase for most

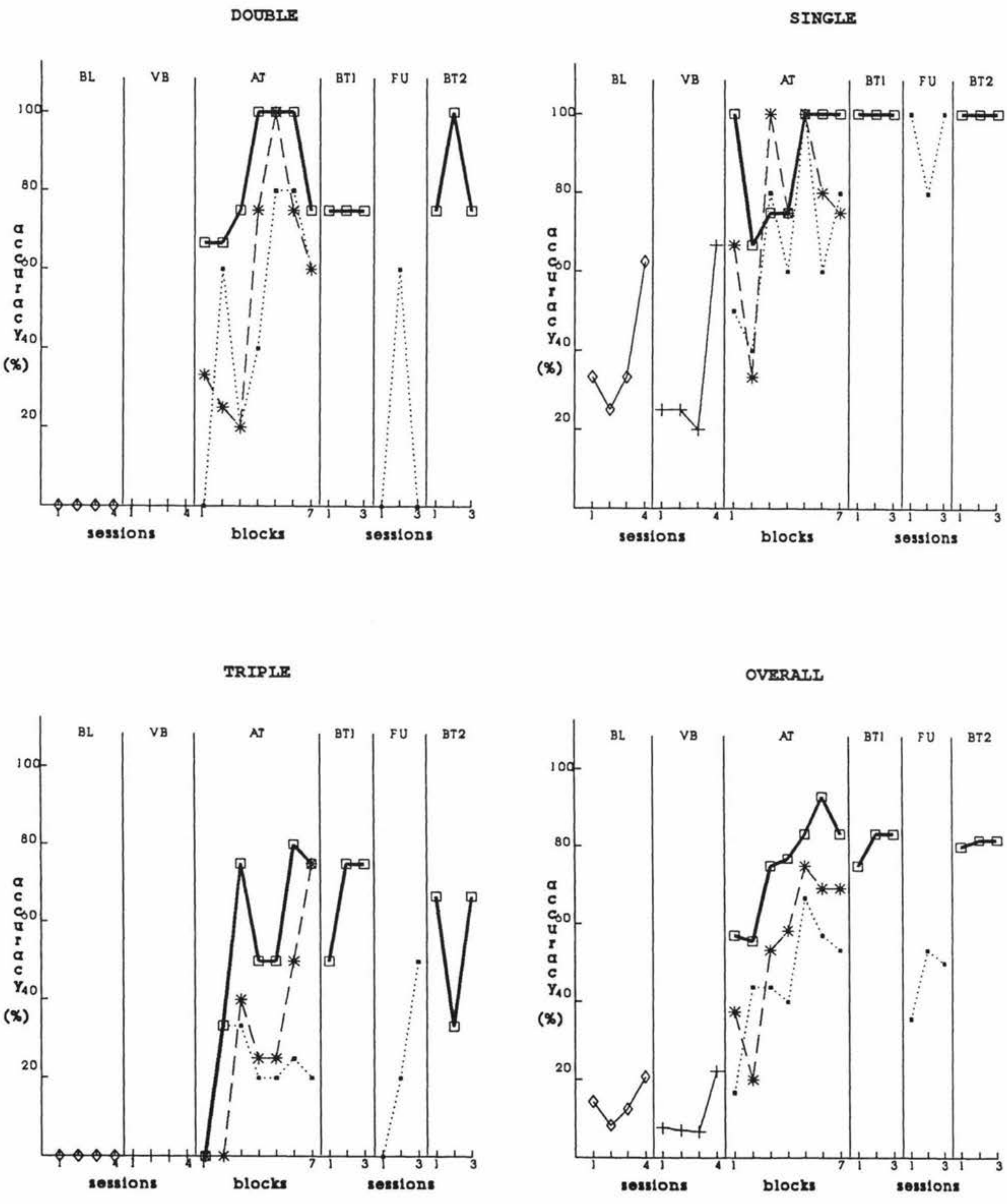
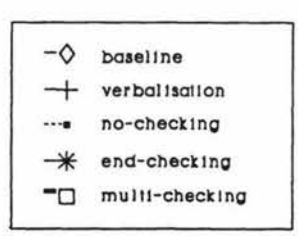


Figure 1: Sue's accuracy scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



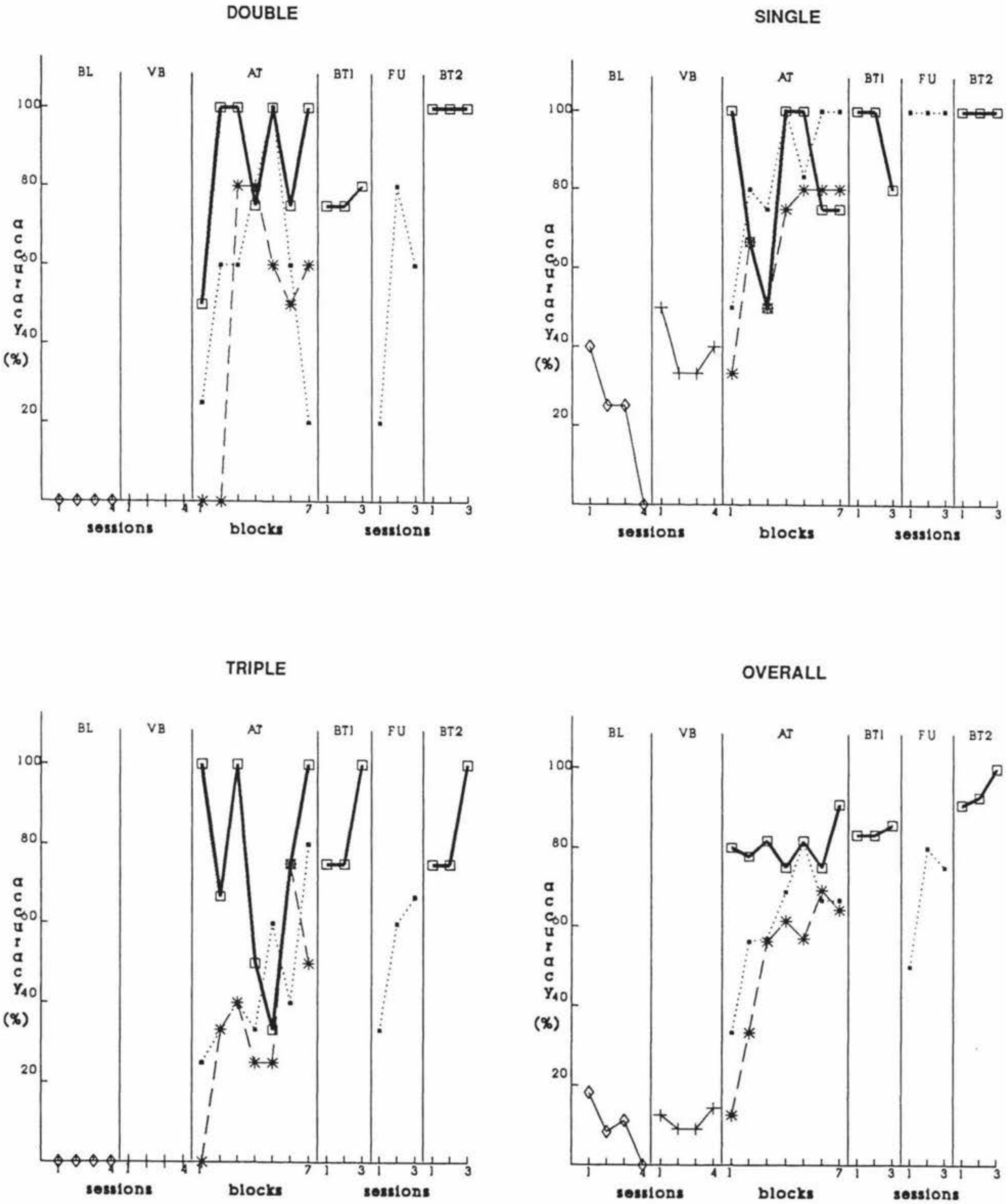
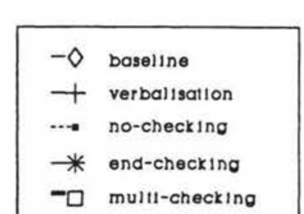


Figure 2: Joe's accuracy scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



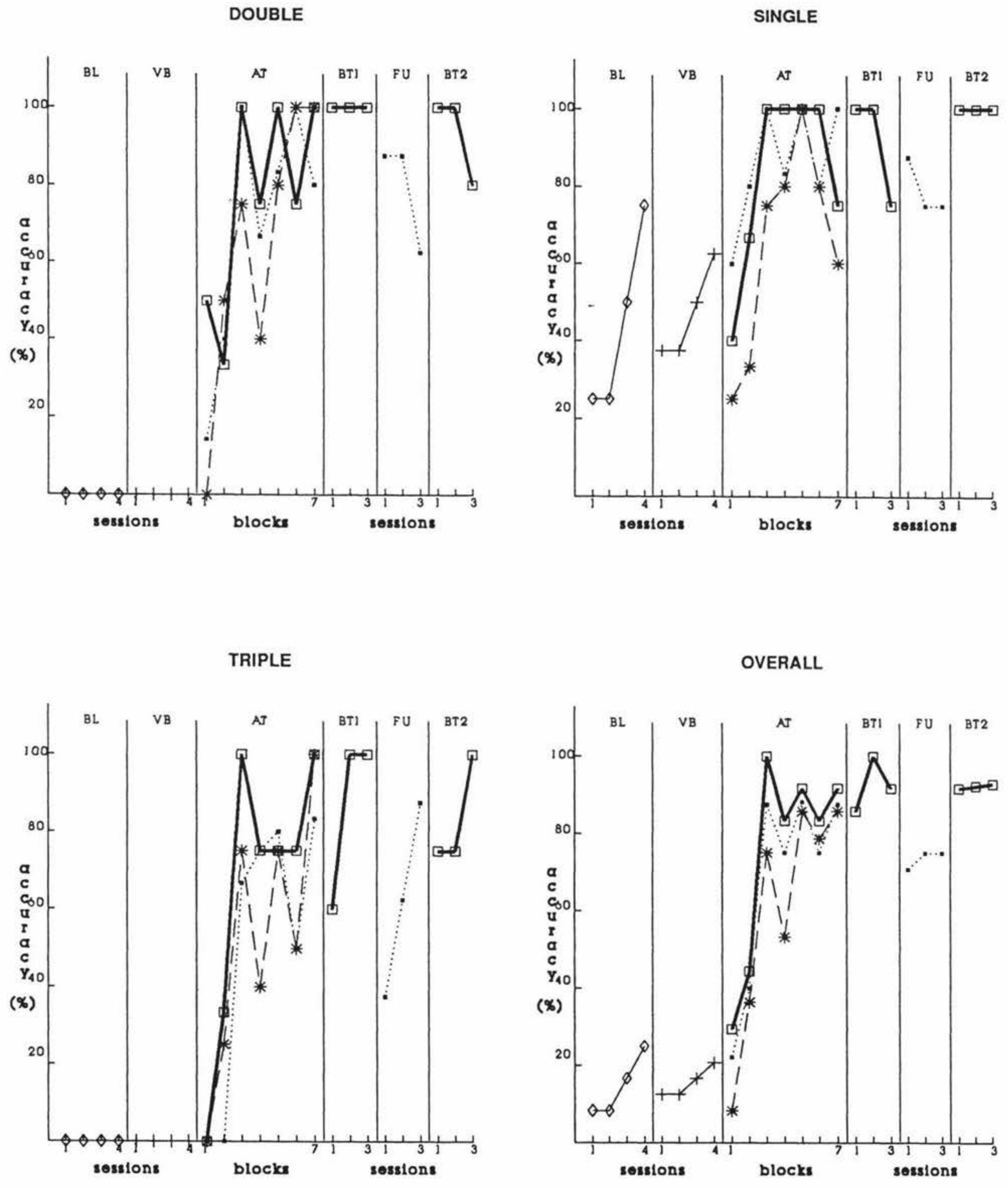
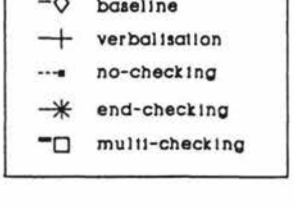


Figure 3: Vera's accuracy scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



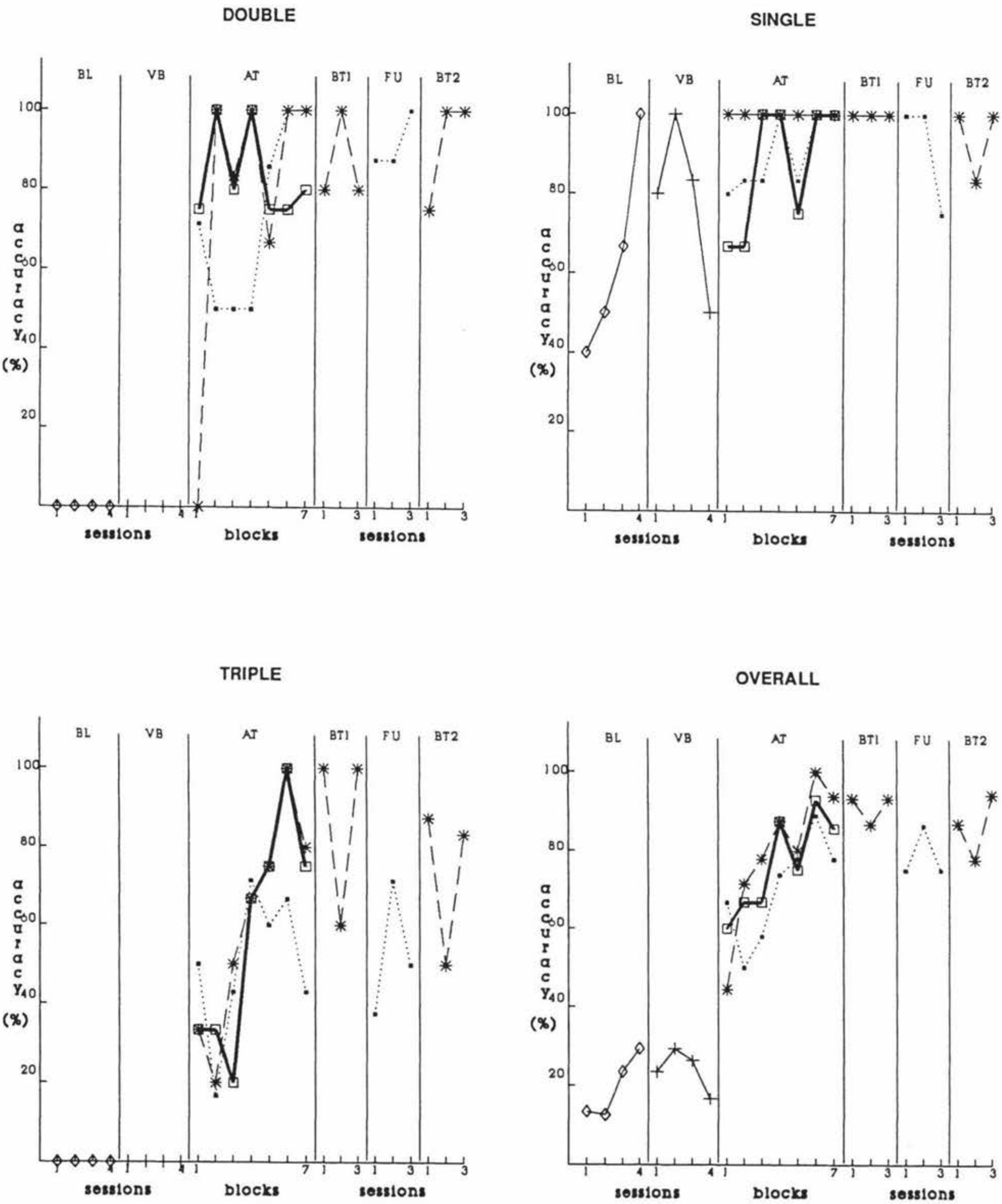
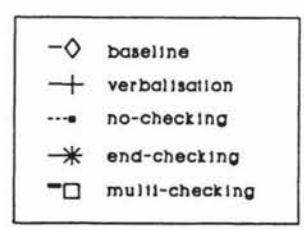


Figure 4: Keith's accuracy scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



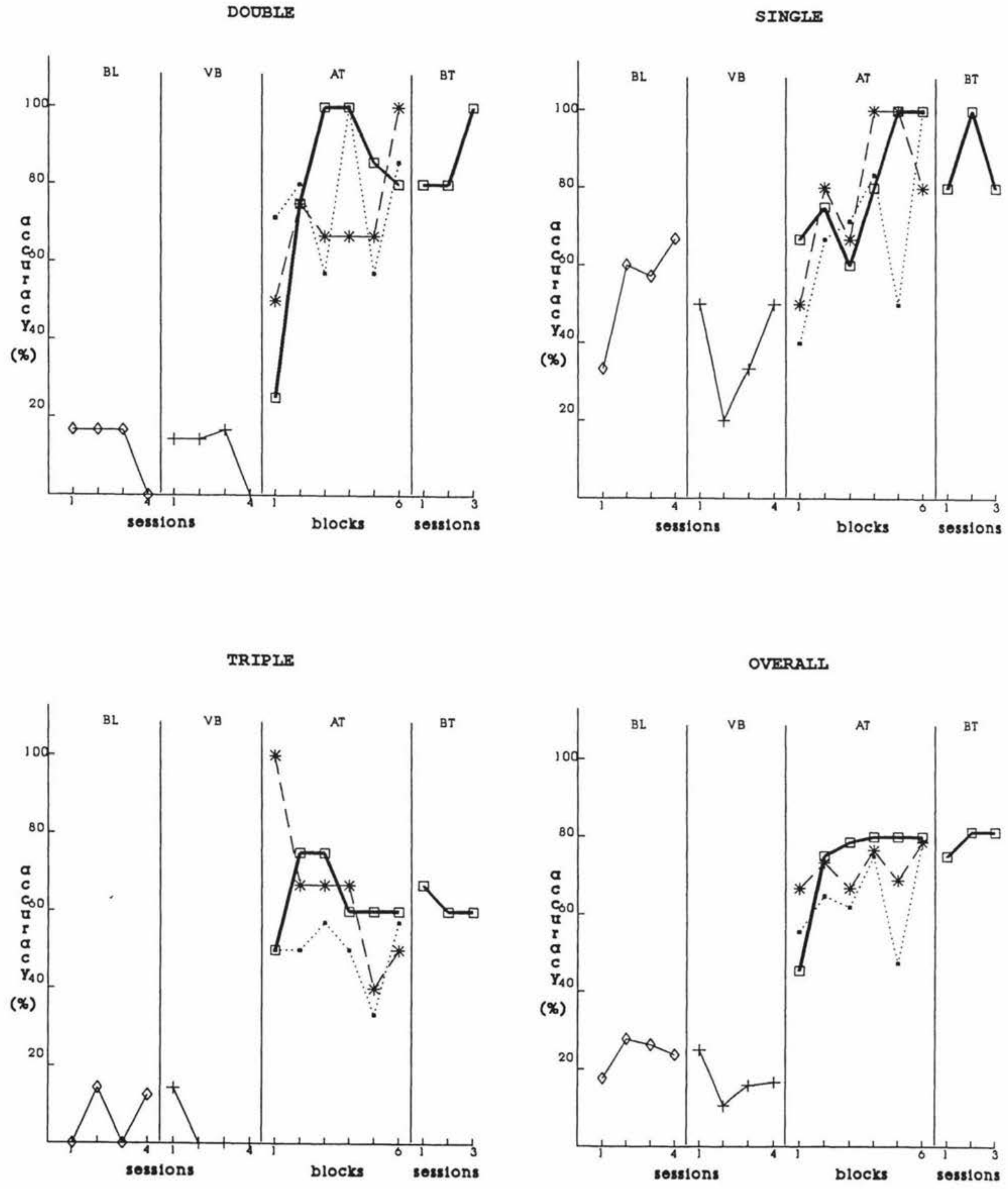
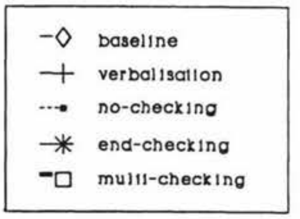


Figure 5: Mary's accuracy scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), and best treatment (BT)



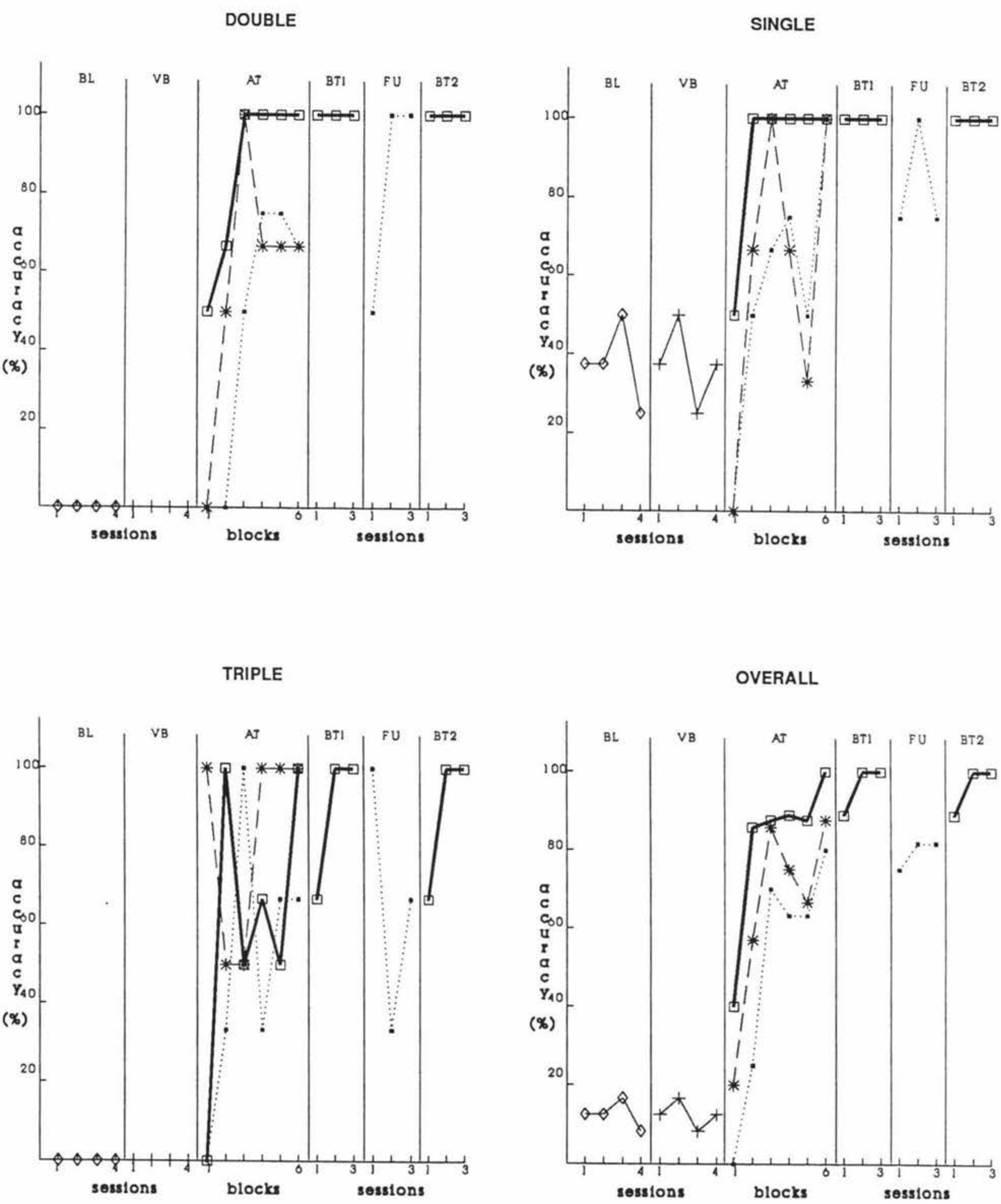
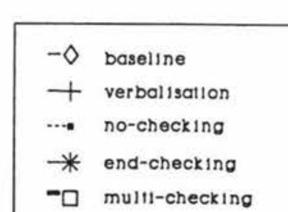


Figure 6: Jane's accuracy scores for each division problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



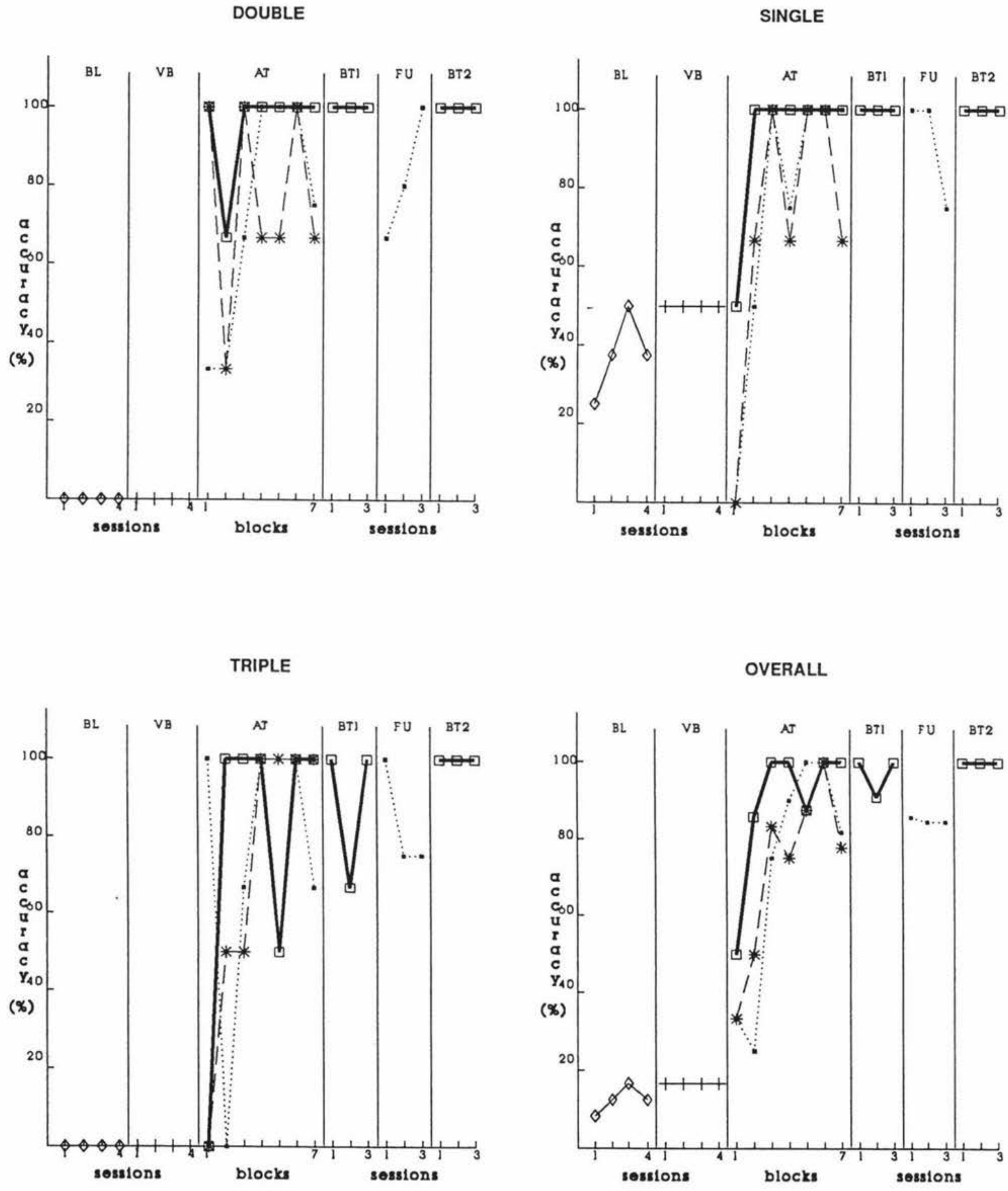
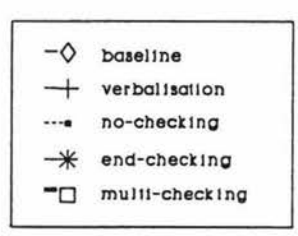


Figure 7: Joan's accuracy for each division problem cluster, and overall problems during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



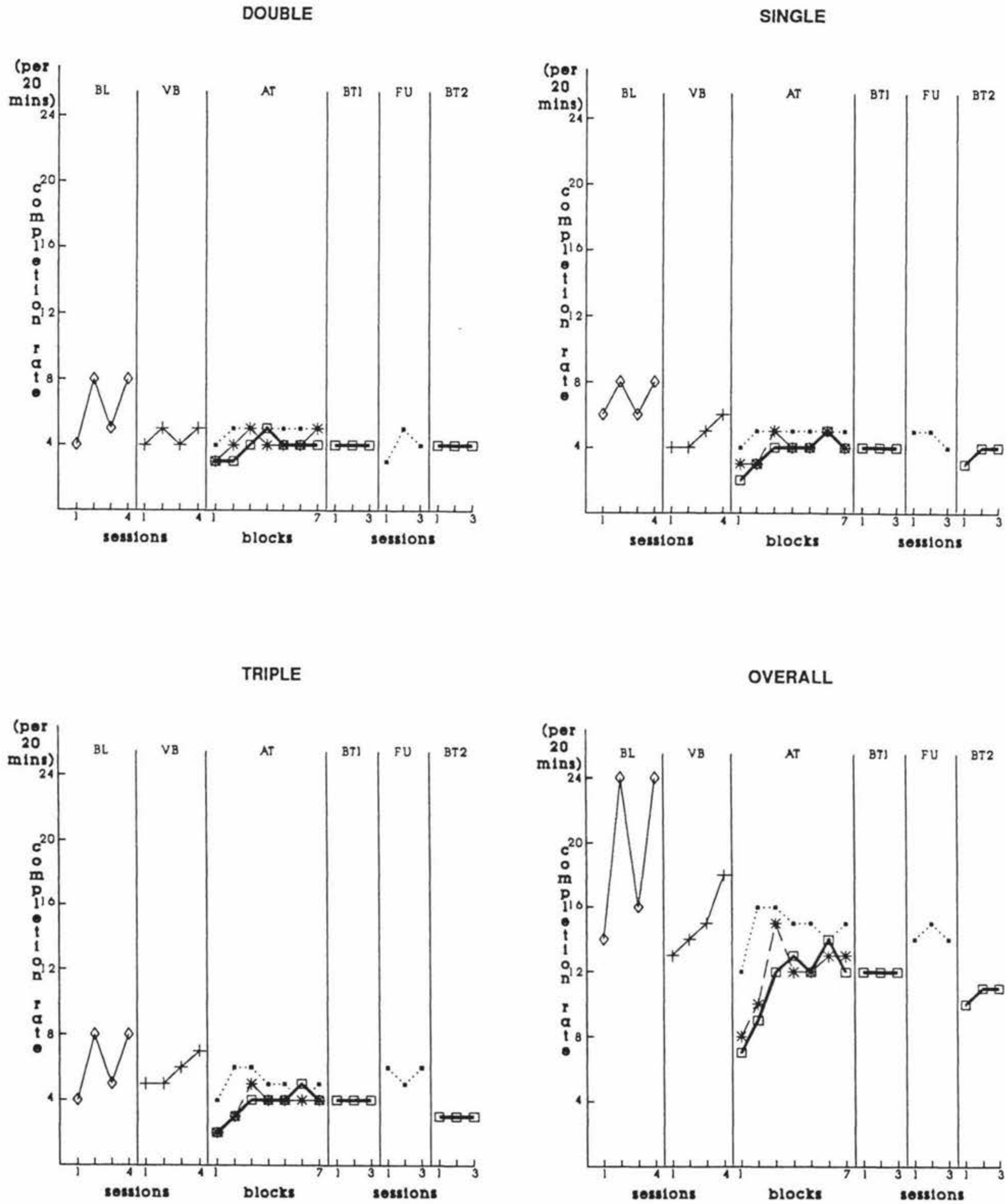
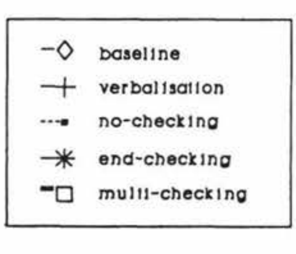


Figure 8: Sue's completion rate scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



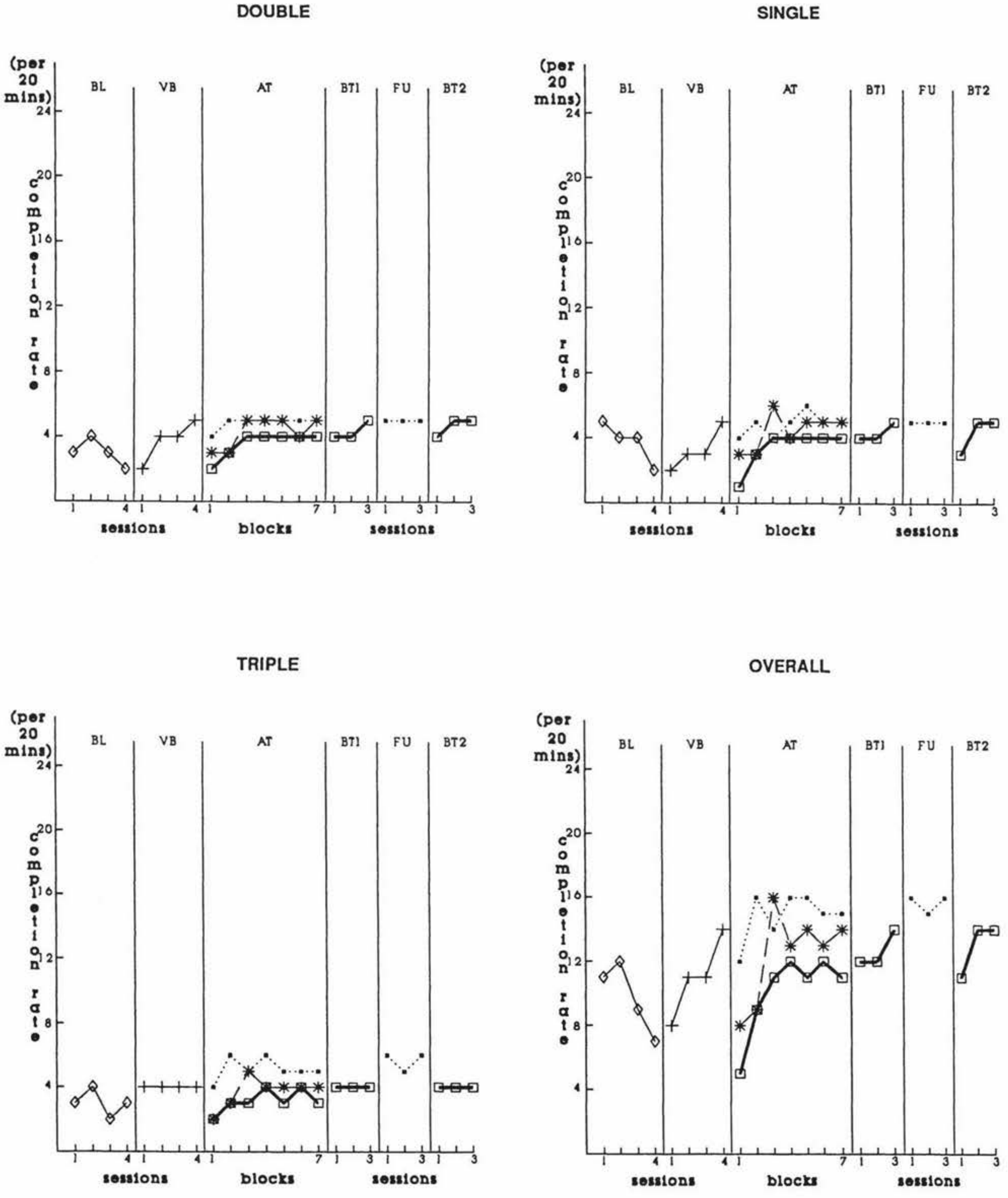
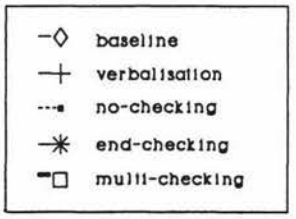


Figure 9: Joe's completion rate scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



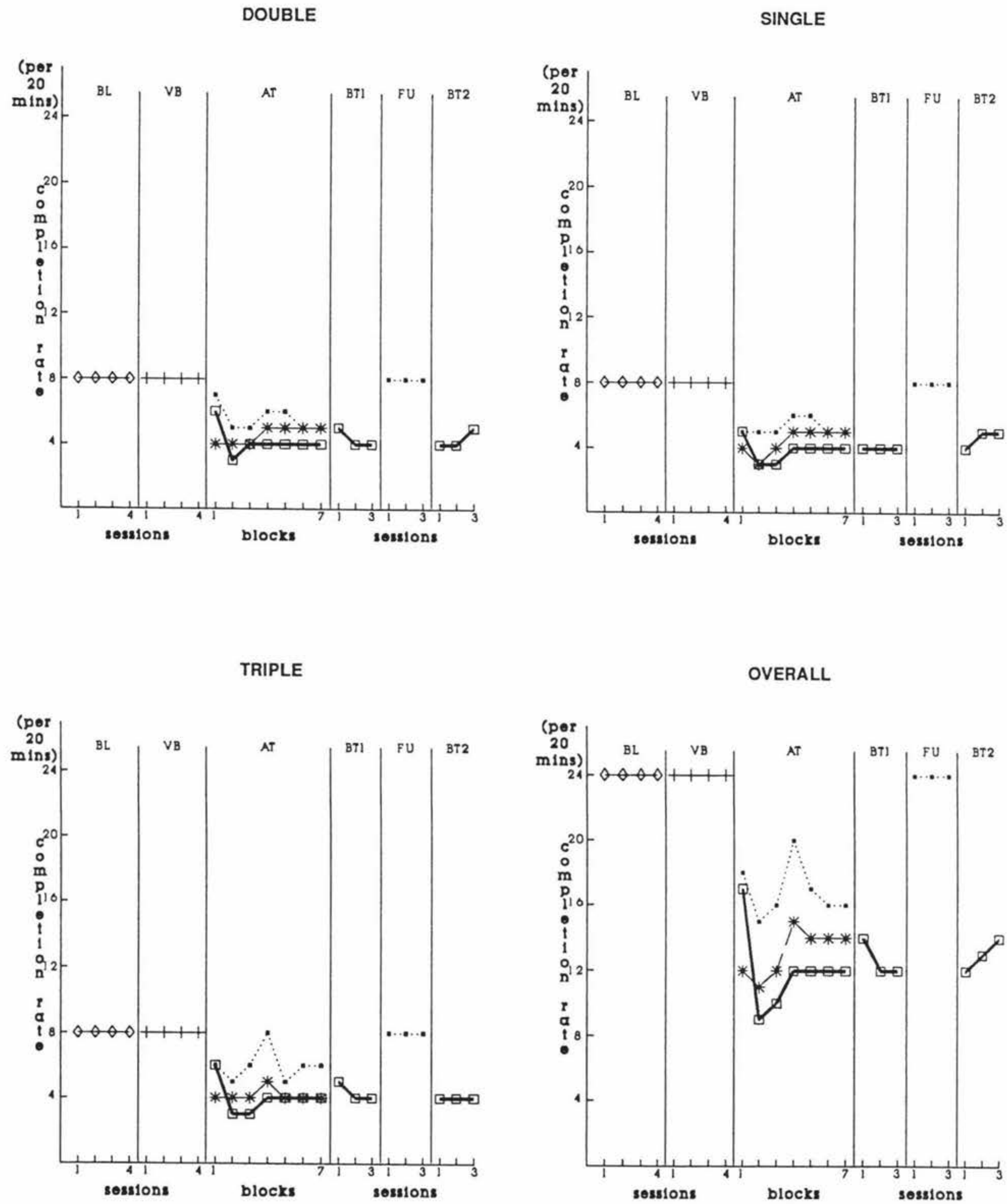
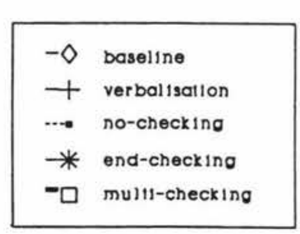


Figure 10: Vera's completion rate scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



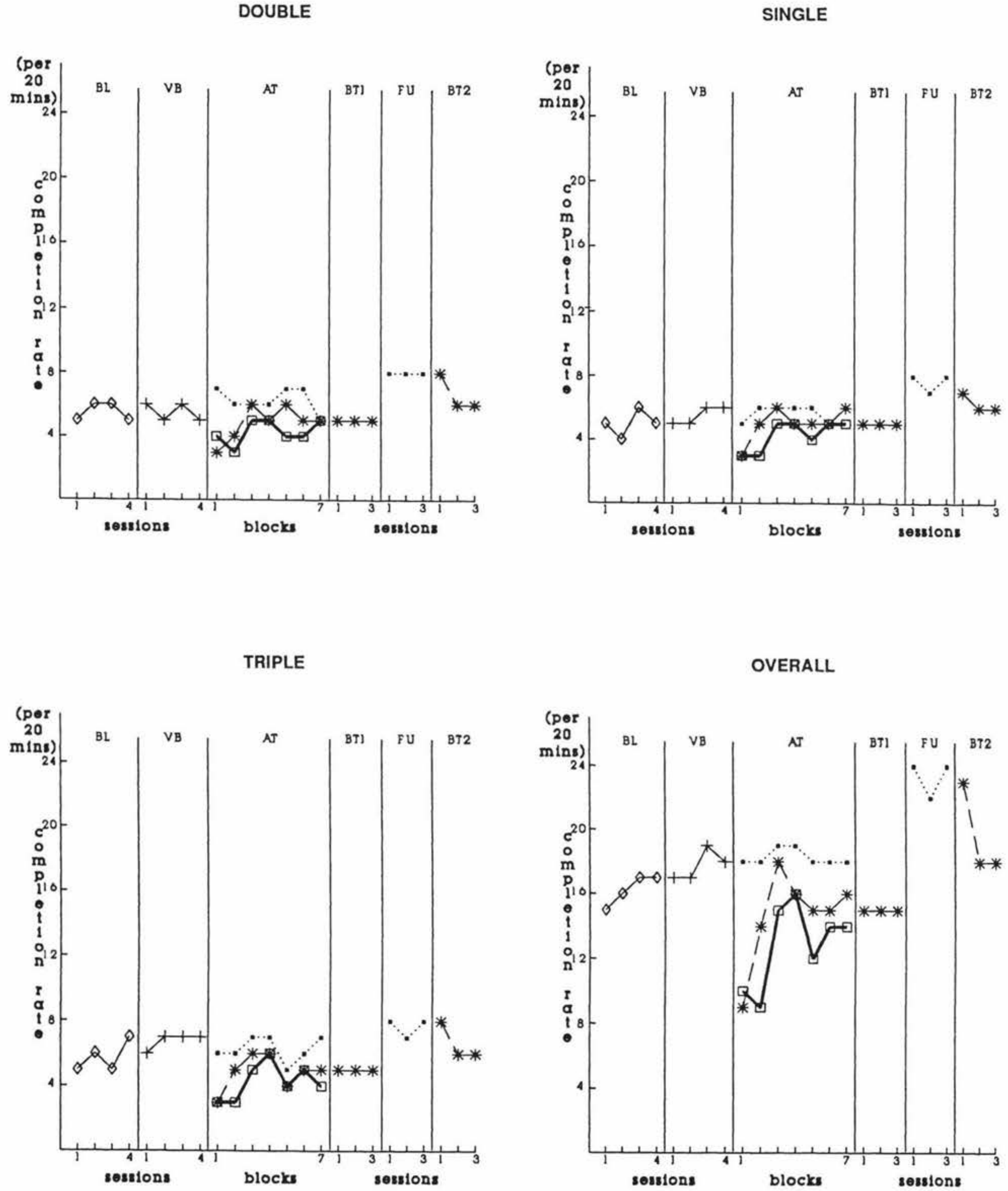
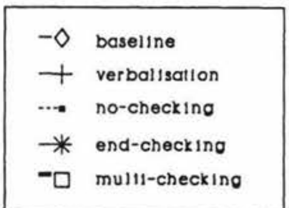


Figure 11: Keith's completion rate scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



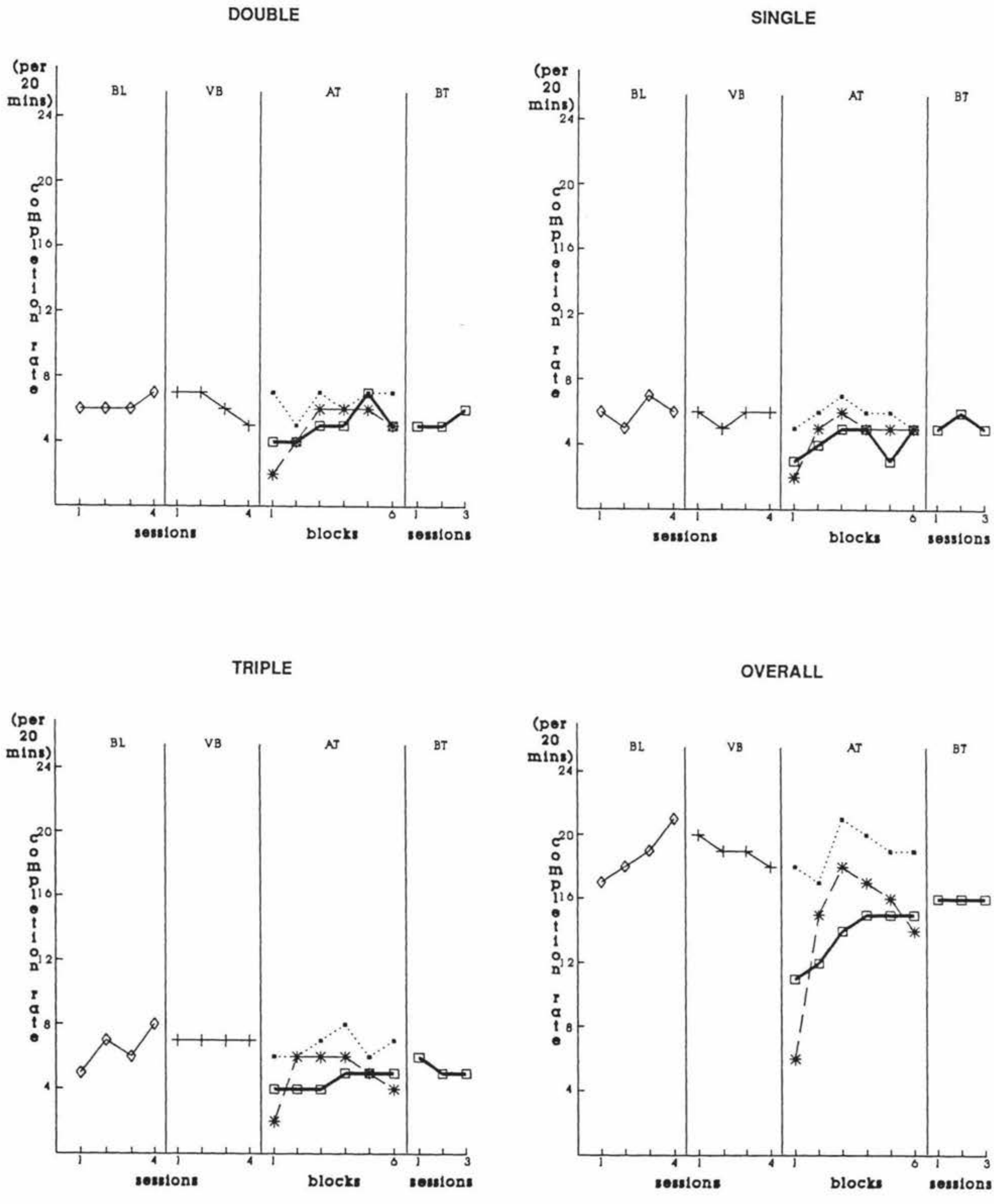
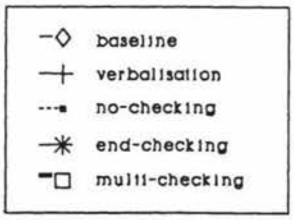


Figure 12: Mary's completion rate scores for each multiplication problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), and best treatment (BT)



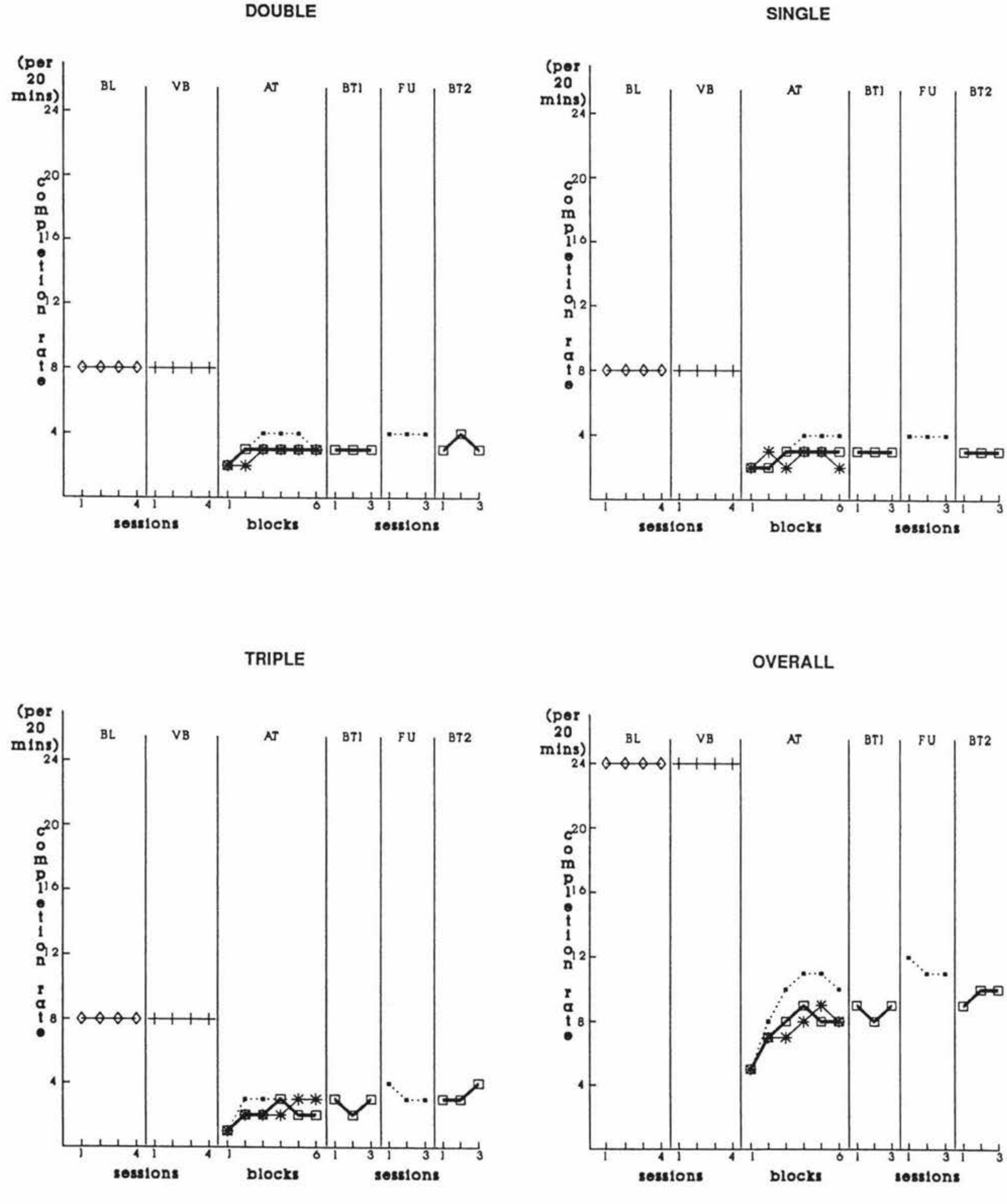
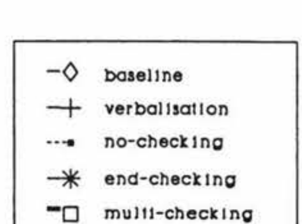


Figure 13: Jane's completion rate scores for each division problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)



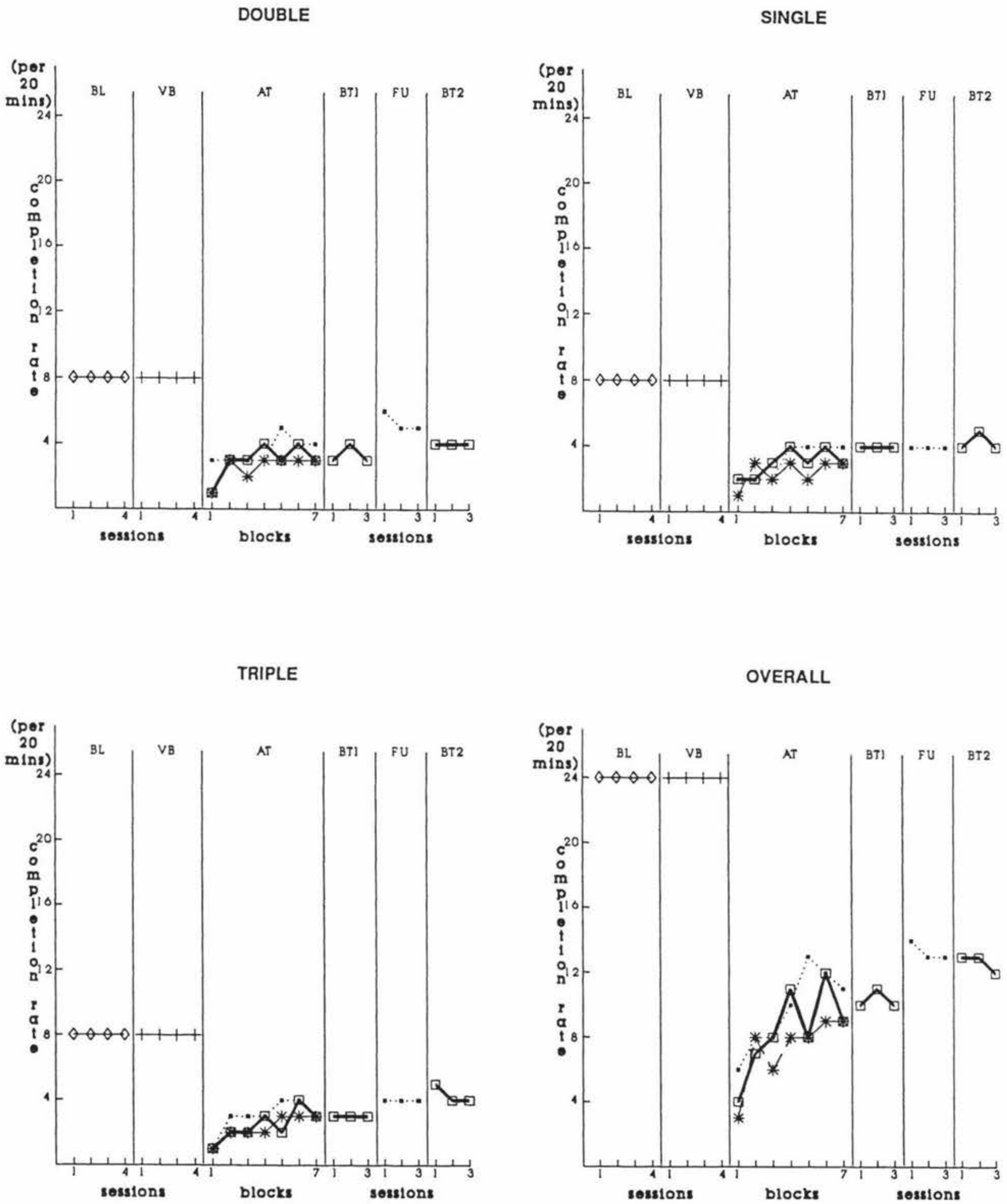


Figure 14: Joan's completion rate scores for each division problem cluster, and overall during baseline (BL), verbalisation (VB), alternating treatments (AT), best treatment 1 (BT1), 6-week follow-up (FU), and best treatment 2 (BT2)

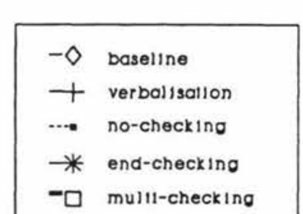


Table 4

Sue's mean scores for accuracy, completion rate, and error rate on each multiplication problem cluster, and overall across all phases of the experiment

	baseline	verbalisation	alternating treatments: no checking	alternating treatments: end checking	alternating treatments: multi checking	best treatment 1	follow-up	best treatment 2
ACCURACY (percentage)								
double	0.0	0.0	48.6	55.5	83.3	75.0	20.0	83.3
single	38.5	34.2	67.1	75.7	88.1	100.0	93.3	100.0
triple	0.0	0.0	21.7	30.7	51.9	66.7	23.3	55.6
overall	14.0	11.0	45.9	54.6	74.9	80.5	46.3	81.2
COMPLETION RATE (per 20 min)								
double	6.3	4.5	4.9	4.1	3.9	4.0	4.0	4.0
single	7.0	4.8	4.9	4.0	3.7	4.0	4.7	3.7
triple	6.3	5.8	5.0	3.7	3.7	4.0	5.7	3.0
overall	19.5	15.0	14.7	11.8	11.3	12.0	14.3	10.7
ERROR RATE (per 20 min)								
double	6.3	4.5	2.4	1.9	0.6	1.0	3.0	0.7
single	4.3	3.0	1.6	0.9	0.4	0.0	0.3	0.0
triple	6.3	5.8	3.9	2.4	1.6	1.3	4.3	1.3
overall	16.8	13.3	7.9	5.1	2.6	2.3	7.7	2.0

Table 5

Joe's mean scores for accuracy, completion rate, and error rate on each multiplication problem cluster, and overall across all phases of the experiment

	baseline	verbalisation	alternating treatments: no checking	end checking	multi checking	best treatment 1	follow-up	best treatment 2
ACCURACY (percentage)								
double	0.0	0.0	57.9	47.1	85.7	76.7	53.3	100.0
single	22.5	39.2	84.0	66.4	81.0	93.3	100.0	100.0
triple	0.0	0.0	44.5	35.5	75.0	83.3	53.3	83.3
overall	9.4	11.3	61.5	50.6	80.3	84.1	68.3	94.6
COMPLETION RATE (per 20 min)								
double	3.0	3.8	4.9	4.3	3.6	4.3	5.0	4.7
single	3.8	3.3	4.9	4.4	3.4	4.3	5.0	4.3
triple	3.0	4.0	5.1	3.7	3.1	4.0	5.7	4.0
overall	9.8	11.0	14.9	12.4	10.1	12.7	15.7	13.0
ERROR RATE (per 20 min)								
double	3.0	3.8	2.0	2.0	0.4	1.0	2.3	0.0
single	2.8	2.0	0.7	1.4	0.7	0.3	0.0	0.0
triple	3.0	4.0	2.9	2.3	0.9	0.7	2.7	0.7
overall	8.8	9.8	5.6	5.7	2.0	2.0	5.0	0.7

Table 6

Vera's mean scores for accuracy, completion rate, and error rate on each multiplication problem cluster, and overall across all phases of the experiment

	baseline	verbalisation	alternating treatments: no checking	end checking	multi checking	best treatment 1	follow-up	best treatment 2
ACCURACY (percentage)								
double	0.0	0.0	69.2	63.6	76.2	100.0	79.2	93.3
single	43.8	46.9	86.2	64.8	83.1	91.7	79.2	100.0
triple	0.0	0.0	50.7	52.1	65.5	86.7	62.5	83.3
overall	14.6	15.6	67.9	60.4	74.8	92.4	73.6	92.3
COMPLETION RATE (per 20 min)								
double	8.0	8.0	5.6	4.6	4.1	4.3	6.3	8.0
single	8.0	8.0	5.3	4.4	3.9	4.0	8.0	4.7
triple	8.0	8.0	6.0	4.1	4.0	4.3	8.0	4.0
overall	24.0	24.0	16.9	13.1	12.0	12.7	24.0	13.0
ERROR RATE (per 20 min)								
double	8.0	8.0	1.9	1.6	1.0	0.0	1.7	0.3
single	4.5	4.3	0.7	1.4	0.7	0.3	1.7	0.0
triple	8.0	8.0	2.9	2.0	1.6	0.7	3.0	0.7
overall	20.5	20.3	5.4	5.0	3.3	1.0	6.3	1.0

Table 7

Keith's mean scores for accuracy, completion rate, and error rate on each multiplication problem cluster, and overall across all phases of the experiment

	baseline	verbalisation	alternating treatments: no checking	alternating treatments: end checking	alternating treatments: multi checking	best treatment 1	follow-up	best treatment 2
ACCURACY (percentage)								
double	0.0	0.0	72.4	78.6	83.6	86.7	91.7	91.7
single	64.2	78.3	90.0	100.0	86.9	100.0	91.7	94.4
triple	0.0	0.0	50.1	60.7	57.6	86.7	53.0	73.6
overall	19.7	24.0	70.4	79.3	76.4	91.1	78.8	86.4
COMPLETION RATE (per 20 min)								
double	5.5	5.5	6.3	4.9	4.3	5.0	8.0	6.7
single	5.0	5.5	5.7	5.0	4.3	5.0	7.7	6.3
triple	5.8	6.8	6.3	4.9	4.3	5.0	7.7	6.7
overall	16.3	17.8	18.3	14.7	12.9	15.0	23.3	19.7
ERROR RATE (per 20 min)								
double	5.5	5.5	1.7	0.9	0.7	0.7	0.7	0.7
single	1.8	1.3	0.6	0.0	0.4	0.0	0.7	0.3
triple	5.8	6.8	3.1	1.9	1.7	0.7	3.7	1.7
overall	13.0	13.5	5.4	2.7	2.9	1.3	5.0	2.7

Table 8

Mary's mean scores for accuracy, completion rate, and error rate on each multiplication problem cluster, and overall across all phases of the experiment

			alternating treatments			
	baseline	verbalisation	no checking	end checking	multi checking	best treatment
ACCURACY (percentage)						
double	12.5	11.3	75.2	70.9	77.6	86.7
single	54.3	38.3	68.6	79.5	80.3	86.7
triple	6.7	3.6	49.6	65.0	63.3	62.2
overall	23.8	17.0	63.9	71.8	73.2	79.2
COMPLETION RATE (per 20 min)						
double	6.3	6.3	6.5	4.8	5.0	5.3
single	6.0	5.8	5.8	4.7	4.2	5.3
triple	6.5	7.0	6.7	4.8	4.5	5.3
overall	18.8	19.0	19.0	14.3	13.7	16.0
ERROR RATE (per 20 min)						
double	5.5	5.5	1.7	1.3	1.0	0.7
single	2.8	3.5	1.8	0.8	0.8	0.7
triple	6.0	6.8	3.3	1.8	1.7	2.0
overall	14.3	15.8	6.8	4.0	3.5	3.3

Table 9

Jane's mean scores for accuracy, completion rate, and error rate on each division problem cluster, and overall across all phases of the experiment

	baseline	verbalisation	alternating treatments: no checking	alternating treatments: end checking	alternating treatments: multi checking	best treatment 1	follow-up	best treatment 2
ACCURACY (percentage)								
double	0.0	0.0	44.5	58.4	86.1	100.0	83.3	100.0.
single	37.5	37.5	57.0	61.1	91.7	100.0	83.3	100.0
triple	0.0	0.0	50.0	83.3	61.1	88.9	66.7	88.9
overall	12.5	12.5	50.4	65.3	81.6	96.3	79.5	96.3
COMPLETION RATE (per 20 min)								
double	8.0	8.0	3.3	2.7	2.8	3.0	4.0	3.3
single	8.0	8.0	3.2	2.5	2.7	3.0	4.0	3.0
triple	8.0	8.0	2.7	2.2	2.0	2.7	3.3	3.3
overall	24.0	24.0	9.2	7.3	7.5	8.7	11.3	9.7
ERROR RATE (per 20 min)								
double	8.0	8.0	1.7	1.0	0.3	0.0	0.7	0.0
single	5.0	5.0	1.2	1.0	0.2	0.0	0.7	0.0
triple	8.0	8.0	1.2	0.3	0.7	0.3	1.0	0.3
overall	21.0	21.0	4.0	2.3	1.2	0.3	2.3	0.3

Table 10

Joan's mean scores for accuracy, completion rate, and error rate on each division problem cluster, and overall across all phases of the experiment

			alternating treatments					
	baseline	verbalisation	no checking	end checking	multi checking	best treatment 1	follow-up	best treatment 2
ACCURACY (percentage)								
double	0.0	0.0	72.6	76.2	95.2	100.0	82.2	100.0
single	37.5	50.0	75.0	71.4	92.9	100.0	91.7	100.0
triple	0.0	0.0	76.2	71.4	78.6	88.9	83.3	100.0
overall	12.5	16.7	72.2	72.4	89.0	97.0	85.0	100.0
COMPLETION RATE (per 20 min)								
double	8.0	8.0	3.6	2.6	3.0	3.3	5.3	4.0
single	8.0	8.0	3.1	2.4	3.0	4.0	4.0	4.3
triple	8.0	8.0	3.0	2.3	2.4	3.0	4.0	4.3
overall	24.0	24.0	9.7	7.3	8.4	10.3	13.3	12.7
ERROR RATE (per 20 min)								
double	8.0	8.0	0.9	0.7	0.1	0.0	1.0	0.0
single	5.0	4.0	0.6	0.6	0.1	0.0	0.3	0.0
triple	8.0	8.0	0.7	0.4	0.3	0.3	0.7	0.0
overall	21.0	20.0	2.1	1.7	0.6	0.3	2.0	0.0

multiplication students (mean range = 22.5% - 64.2%), with variability being moderate to large. Accuracy scores on sdd problems were at similar levels, had much less variability and showed no consistent trends across the phase (mean for both students = 37.5%). On tdm/tdd problems, all students scored 0% in all sessions except Mary who scored above 10% on two occasions (mean = 6.7%).

All multiplication students had low levels of overall accuracy (mean range = 9.4% - 23.8%), with moderate variability and no consistent trends across subjects. For division students, the accuracy levels were low (both having a mean of 12.5%) and showed little variability or trend.

Each type of problem cluster constituted one third of the 24 problems in each worksheet for each session, so the number completed for each problem cluster was small (range = 1 - 8). As students were given 20 minutes to complete the worksheet, they finished at different points in the worksheet. This generally resulted in each student completing unequal but comparable numbers of each problem type. Although these numbers varied across students (mean range = 3 - 8), each student, tended to show fairly constant rates across the phase. The data of Vera, Jane, and Joan were very stable because they completed all problems each day, although, except for some sdm/sdd problems, incorrectly.

Overall completion rates showed no consistency in either levels or trends across students (mean range = 9.8 - 24.0). Only Sue showed large variability across the phase.

Mean error rates were generally lower for the sdm/sdd problems (mean range = 1.8 - 5.0) than on ddm/ddd problems (mean range = 3.0 - 8.0) and tdm/tdd problems (mean range = 3.0 - 8.0). Overall mean error rates were high (mean range = 8.8 - 21.0). Students who had the highest completion rates also had the highest error rates.

VERBALISATION (VB)

In the verbalisation phase, accuracy was again 0% in the multidigit multiplier/divisor problems for all but one student. Levels and variability of the accuracy scores on the sdm/sdd problems were usually comparable to those in the BL phase. Overall accuracy levels and variability were also comparable to those shown in the BL phase.

Completion rate levels and variability for each student were also similar to those shown in the BL phase, with mean completion rates for the problem clusters ranging from 3.3 to 8.0, and mean overall completion rates ranging from 11.0 to 24.0.

Mean error rates on the problem clusters and overall were also generally comparable to those in the BL phase. Only for one student (Sue) were mean error rates slightly lower than those in the BL phase for each of the problem clusters and overall.

ALTERNATING TREATMENTS (AT)

Problem cluster accuracy

During the AT phase, accuracy increased rapidly and substantially under all treatment procedures for all students on all problem clusters. In general, for each student, each procedure produced an increasing trend, although with marked reversals at times. This trend usually levelled off at high but differing levels of accuracy and with marked session by session fluctuations for some. For some students, under some procedures, further increases or decreases occurred towards the end of the phase. For all students except Keith, mc produced the highest accuracy, or equal highest on most blocks of the phase. However, overlapping of data points occurred under the three procedures for all students on all three problem clusters and also overall (though to a much lesser extent). Thus statistical analyses were conducted to complement visual inspection of the data.

Autocorrelations (of lag 1) were computed for each student for each problem cluster and overall, to test for serial dependency in the scores. This was done because significant autocorrelations can greatly bias *t* - and *F* -tests (Kazdin, 1982). No autocorrelation was statistically significant. Accuracy between the various treatment procedures for each problem cluster and overall were then compared for each subject using F-tests which, if significant were followed by correlated *t* -tests. Table 11 shows the results of the *F* -tests and two-tailed correlated *t* -tests for each subject for all problem clusters and overall. All *t* -tests are presented whether significant or not except where the *F* -test proved insignificant. The level of significance for all tests was set at $p < 0.05$.

Mean accuracy scores on ddm/ddd problems showed substantial increases over those in the VB phase for all students under all treatment conditions (and on occasions reached 100%). Mean accuracy scores ranged from 44.5% to 75.2% under nc, 47.1% to 78.6% under ec and 76.2% to 95.2% under mc. Although there was often considerable overlap and variability in the accuracy scores under the three procedures (especially ec and nc) across the phase, mc produced higher mean accuracy scores than ec and nc for all subjects, including Keith. Generally there were rapid and sharp increasing trends in the accuracy scores under mc, which levelled off, although with some variability, at high mean levels with division students scoring even higher mean levels. Mc scores were equal to

Table 11

Summary of comparisons between treatments in the alternating phase for each student on the accuracy scores for each multiplication/division problem cluster, and overall, using *F* -tests and two-tailed correlated *t* -tests

Subjects	accuracy measure	*F* -value and *p*	Comparisons	*t* -value	*p* (less than)	significantly most effective treatment
Sue	double	9.24 (p < 0.004)	mc vs nc	4.059	0.01	mc
			mc vs ec	3.253	0.01	mc
			ec vs nc	0.805	n.s.	-
	single	4.63 (p < 0.032)	mc vs nc	3.025	0.02	mc
			mc vs ec	1.788	n.s.	-
			ec vs nc	1.237	n.s.	-
	triple	6.89 (p < 0.010)	mc vs nc	3.617	0.01	mc
			mc vs ec	2.534	0.05	mc
			ec vs nc	1.083	n.s.	-
	overall	22.36 (p < 0.000)	mc vs nc	6.518	0.001	mc
			mc vs ec	4.554	0.001	mc
			ec vs nc	1.965	n.s.	-
Joe	double	5.40 (p < 0.021)	mc vs nc	2.299	0.05	mc
			mc vs ec	3.183	0.01	mc
			ec vs nc	0.884	n.s.	-
	single	2.44 (n.s.)	-	-	-	-
	triple	6.40 (p < 0.013)	mc vs nc	2.633	0.05	-
			mc vs ec	3.414	0.01	-
			ec vs nc	0.781	n.s.	-
	overall	11.94 (p < 0.001)	mc vs nc	3.066	0.01	-
			mc vs ec	4.830	0.001	-
			ec vs nc	1.764	n.s.	-
Vera	double	1.15 (n.s.)	-	-	-	-
	single	6.93 (p < 0.032)	mc vs nc	0.496	n.s.	-
			mc vs ec	2.948	0.020	mc
			ec vs nc	3.444	0.010	nc
	triple	3.25 (n.s.)	-	-	-	-
	overall	10.89 (p < 0.002)	mc vs nc	2.240	0.050	mc
			mc vs ec	4.665	0.001	mc
			ec vs nc	2.425	0.050	nc

Subjects	accuracy measure	F -value and p	Comparisons	t -value	p (less than)	significantly most effective treatment
Keith	double	0.33 (n.s.)	-	-	-	-
	single	4.05 (n.s.)	mc vs nc	0.641	n.s.	-
			mc vs ec	2.722	0.02	ec
			ec vs nc	2.082	n.s.	-
	triple	1.18 (n.s.)	-	-	-	-
	overall	2.27 (n.s.)	-	-	-	-
Mary	double	0.21 (n.s.)	-	-	-	-
	single	4.05 (n.s.)	-	-	-	-
	triple	1.18 (n.s.)	-	-	-	-
	overall	2.27 (n.s.)	-	-	-	-
Jane	double	11.83 ($p < 0.002$)	mc vs nc	4.776	0.001	mc
			mc vs ec	3.183	0.01	mc
			ec vs nc	1.593	n.s.	-
	single	8.98 ($p < 0.006$)	mc vs nc	3.882	0.01	mc
			mc vs ec	3.416	0.01	mc
			ec vs nc	0.466	n.s.	-
	triple	1.15 (n.s.)	-	-	-	-
	overall	19.15 ($p < 0.000$)	mc vs nc	6.187	0.001	mc
			mc vs ec	3.223	0.01	mc
			ec vs nc	2.965	0.02	ec
Joan	double	2.79 (n.s.)	-	-	-	-
	single	4.50 ($p < 0.035$)	mc vs nc	2.334	0.05	mc
			mc vs ec	2.799	0.02	mc
			ec vs nc	0.465	n.s.	-
	triple	0.07 (n.s.)	-	-	-	-
	overall	4.45 ($p < 0.036$)	mc vs nc	2.603	0.05	mc
			mc vs ec	2.564	0.05	mc
			ec vs nc	0.040	n.s.	-

or higher than ec and nc scores on most blocks for three of the multiplication students (Sue, Joe and Vera) and for both division students. The division students obtained 100% accuracy under mc on the last 4 or 5 blocks. For Keith, ec was best or equal best for most blocks and for Mary no one procedure was best on most blocks. Statistical analyses revealed that the only significant differences were that mc was more effective than both ec and nc for Sue, Joe, and Jane.

On the sdm/sdd problems accuracy increased substantially over VB phase levels, under all three treatment procedures. Mean accuracy scores ranged from 57.0% to 90.0% under nc, 61.1% to 100.0% under ec, and 80.3% to 92.9% under mc. Scores for the three procedures overlapped on several occasions for all students throughout this phase, but the same procedure that produced the highest or equal highest scores on most blocks for the ddm/ddd problems was the highest or equal highest again in most blocks for sdm problems for Sue and Keith and for sdd problems for Jane and Joan. No treatment was best or equal best on most blocks for Joe and Mary but nc was for Vera. Statistical analyses on the multiplication students' data revealed only the following significant differences: for Sue, mc was more effective than nc, for Keith ec was more effective than mc and for Vera, mc and nc were each more effective than ec. With both division students, the only significant differences were that mc was more effective than both ec and nc.

Generalisation of treatment effects was also found with the tdm/tdd problems where accuracy scores increased substantially over those in the previous phase, under all three procedures. Mean accuracy scores ranged from 21.7% to 76.2% under nc, 30.7% to 83.3% under ec, and 51.9% to 78.6% under mc. Accuracy scores for this cluster were generally lower than for the other two problem clusters. Although there was usually considerable variability and overlap of data points across the phase mc was best or equal best in most blocks for all students except Keith and Jane where ec was best or equal best on most blocks. The only significant effects were that mc was more effective than both ec and nc for Sue and Joe.

Because of the sometimes moderate to large variability, number of reversals and overlapping of data points, it was difficult to determine a procedure which was second most effective on most blocks for most students, on the ddm/ddd problems.

Overall accuracy

Overall accuracy scores showed the differential effectiveness of the three procedures more clearly than the accuracy scores on the individual problem clusters. Under all three procedures all students showed rapid and large increases in overall accuracy from those shown in the VB phase. Mean overall accuracy scores ranged from 45.9% to 72.2% under nc, 50.6% to 79.3% under ec, and 73.2% to 89.0% under mc. Mc produced the greatest improvement in mean overall accuracy for all students except Keith with division students having the highest mean overall accuracy scores. Ec produced the highest mean score for Keith.

Under mc, all multiplication students, showed rapid and sharp increases reaching high mean levels of accuracy. Similar patterns occurred with the scores of the division students, who reached even higher levels.

The most effective procedure in nearly all blocks throughout the AT phase was mc for all students except Keith, for whom it was ec. Clear experimental control by mc was shown in four students (Sue, Joe, Vera and Jane) who had mc scores higher than ec and nc scores on all blocks. Although for Keith, ec scores had a slightly higher mean overall accuracy than that under mc, and were greater than those under mc on most blocks, the differences were fairly slight. Compared to the other students, the differential effect between ec and mc was least with Keith. The only significant differences were that mc was more effective than both nc and ec for Sue, Joe, Vera, Jane and Joan.

Large and rapid increases to high levels of overall accuracy were also seen under ec and nc for most students. Ec produced higher scores than nc in most blocks for three multiplication students (Sue, Keith and Mary) and one division student (Jane). The reverse was true for the other two multiplication students (Joe and Vera) and no consistent differential effect was seen with the other division student (Joan). Thus, no consistent differences between ec and nc occurred across students. For Keith mc was usually more effective than nc in most blocks. The only significant differences were that for Vera nc was more effective than ec and for Jane ec was more effective than nc.

The degree of differential effect of the three procedures on overall accuracy varied across students. For all but Sue, these differences tended to decrease (to varying degrees) during the later part of the phase, especially those between nc and ec, though less clearly for some (e.g., Keith and Mary). For Sue, the differential effect of the procedures was large throughout the phase.

Rate of completion

Figures 8-14 show that a comparable number of problems was completed on each problem cluster, although this number varied across students and treatment procedures. Variability of the rate scores for each problem cluster under each treatment was low. Compared to the VB phase, there was generally a decrease in the number of problems on all problem clusters particularly under ec and mc. The nc levels were generally at or below those in the VB phase.

Although for the individual problem clusters, there was some variability and overlapping, in most blocks the rates of problems completed were generally highest or equal highest under nc, for the multiplication students. For the division students, it was difficult to determine unequivocally which procedure produced the highest rate due to the overlapping of data points. However, for all students, the highest mean completion rate occurred under nc, both overall and for each of the three problem clusters (tables 4 - 10). The lowest mean completion rate was usually under mc for the multiplication students and ec for the division students. Mean completion rates for the problem clusters were low and ranged from 3.1 to 6.7 for the multiplication students and even lower, 2.0 to 3.6, for the division students.

The overall completion rate was more variable than that for the individual problem clusters, with some overlapping of data points across the three conditions. However nc produced the highest scores on all blocks for three students (Vera, Keith, and Mary). For nearly all multiplication students the overall rates of problems completed decreased from that in the VB phase especially under mc (mean range = 10.1 - 13.7) and ec (mean range = 11.8 - 14.7), though for Joe there was a slight increase in the mean rate under ec. However under nc (mean range = 14.7 - 19.0), two students (Joe, Keith) showed increases, one (Mary) showed no change, while the other two showed decreases, over VB phase levels. The overall rate of the division students decreased from VB phase levels under all three treatment conditions with means of 9.2 and 9.7 under nc, means of 7.3 under ec, and means of 7.5 and 8.4 under mc.

While the differential effect of the procedures on completion rate for the individual problem clusters and overall was clearer for some than for others, generally there was an increasing trend followed by some levelling off under all the three treatment procedures.

Error rate

Large decreases in mean error rate occurred for both the problem clusters and overall under each treatment over the VB phase. Mean error rate was lowest under mc (mean range = 0.1 - 1.7).

BEST TREATMENT 1 (BT1)

In the next phase the best treatment in the AT phase for ddm/ddd problems was implemented in all sessions. This was mc for all students except Keith, for whom ec was implemented. A treatment was considered the best treatment if it produced during the AT phase, (a) the highest scores on all blocks (which could include being equal best on one block); failing that, (b) the best (which may include some equal best scores) on most blocks. Sue and Jane met the first criterion which meant that they also met the second one. All others except Mary met the second criterion only. Mary met neither of these criteria. However mc was implemented for her because it produced the highest mean score for ddm problems and also her overall scores on mc met the second criterion.

In general, with the individual problem clusters all students maintained levels of accuracy comparable to those in the later part of the AT phase for that problem type, under the same procedure. Again the division students had higher mean accuracy scores than the multiplication students.

For ddm/ddd problems, accuracy scores were generally high and stable for all students (mean range = 75.0% - 100.0%). For the sdm/sdd problems, the accuracy scores were generally very high and were maintained at 100% for all sessions, except for some students who showed a decrease in the last session (mean range = 86.7% - 100.0%). For tdm/tdd problems, accuracy scores showed no consistent trend across students but were usually lower and more variable than for the other two clusters (mean range = 62.2% - 88.9%).

Overall accuracy scores were generally less variable than accuracy scores for each problem cluster. Throughout the phase all students maintained high overall accuracy levels comparable to those in the later part of the AT phase, under the same procedure. Accuracy scores across the phase differed across students but usually showed low variability and a slight increase which levelled off later. Mean overall accuracy scores were high for multiplication students (mean range = 79.2% - 92.4%) and even higher for division students (mean range = 96.3% - 97.0%).

Completion rates were fairly stable for each problem cluster and overall. These rate levels were usually at or slightly above those in the later part of the AT phase for the same procedure for all students. Completion rates were reasonably similar across individual problem clusters for each student, although this number varied slightly across students. Mean completion rates for the problem clusters ranged from 4.0 to 5.3 for multiplication students and 2.7 to 4.0 for division students. Mean overall completion rates ranged from 12.0 to 16.0 for multiplication students and 8.7 to 10.3 for division students.

Error rates were usually comparable to those in the later part of the AT phase for the same procedure, for each problem cluster and overall (see raw data in Appendix 3).

FOLLOW-UP (FU)

Maintenance data were obtained six weeks after the BT1 phase, for all students except Mary, who withdrew from the experiment. During all sessions, every student chose to use nc when solving their problems. Accuracy decreased from that during the BT1 phase on all problem clusters for nearly all students, especially with the multidigit multiplier/divisor problems where variability was also greater.

On ddm/ddd problems, the nc accuracy scores of Sue and Joe were very variable like those under nc in the AT phase. The nc accuracy scores for Keith and the division students increased over the FU phase reaching 100% by at least the last session. Generally the accuracy levels were at or above those of the later part of the AT phase for nc. Mean accuracy decreased from the BT1 phase levels for all students except Keith (mean range = 20.0% - 91.7%), although they were higher under nc in the AT phase except for Sue and Joe.

Generally the accuracy scores on sdm/sdd problems were very high with low variability (mean range = 79.2% - 100.0%). Although for all students mean accuracy levels decreased slightly compared to the BT1 phase, they were much higher than under nc in the AT phase.

On tdm/tdd problems, all but Joan showed a decrease in accuracy from the previous phase, although by variable amounts and reached levels comparable to those under nc in the later part of the AT phase for that problem type (mean range = 23.3% - 83.3%). Joan's level of accuracy was comparable to that in the previous phase. Variability across the phase was usually high.

The overall accuracy scores decreased from those obtained in the BT1 phase for all students, although by markedly different amounts (mean range = 68.3% - 85%). These scores generally showed low variability across the phase with a levelling off, usually after an increase. Overall accuracy levels were generally comparable to those in the later part of the AT phase under nc.

Completion rates on each problem cluster were fairly stable and generally higher than in the BT1 phase for all students (mean range = 3.3 - 8.0). Completion rates were generally similar across clusters, but with some variation across students. Individual problem cluster completion rates were at or above those of nc in the later part of the AT phase.

Overall completion rate increased over the previous phase although by markedly different amounts across students (mean range = 11.3 - 24.0). These rates were at or above those under nc in the later part of the AT phase and were usually fairly stable.

Mean error rates again increased over the previous phase for overall and the individual problem clusters.

BEST TREATMENT 2 (BT2)

During the last phase, when best treatment was again administered, accuracy scores on each of the problem clusters reached levels comparable (or sometimes higher) to those in the BT1 phase except for Sue and Keith where they showed some drop (although not for all clusters).

On ddm/ddd problems accuracy increased to very high levels (mean range = 83.3% - 100.0%), with division students maintaining 100% on all three sessions. The scores of the multiplication students were moderately variable but with no consistent trend across students. On sdm/sdd problems all students increased their accuracy level and all, except for Keith in one session, maintained 100% accuracy in all three sessions. On tdm problems, accuracy scores were generally lower than on the other two problem clusters (mean range = 55.6% - 83.3%), and were usually of moderate variability. Though the scores of most students were comparable to the BT1 scores, this was not so for Sue and Keith, whose scores were somewhat lower. On tdd problems 100% accuracy was maintained in all sessions except by Jane in the first session (mean range = 88.9% - 100.0%).

Mean overall accuracy scores of all students increased from the FU phase, although by varying amounts (mean range = 81.2% - 100.0%). For all but Keith, mean accuracy scores were comparable to or higher than those in the BT1 phase. Except for Joan, who had 100% in all sessions, accuracy scores increased somewhat over the phase.

While completion rate levels on each problem cluster and overall decreased from those in the FU phase they were generally at or above those in the BT1 phase. A similar number of problems was completed under each problem cluster though this varied across students and treatments. These rate scores were fairly stable and generally showed slight increases over the phase (mean range = 3.0 - 8.0).

Mean error rates dropped again with the reinstatement of the best treatment phase, for all students with nearly all problem clusters. These rates were generally comparable to the means in the BT1 phase. Overall mean error rates were lower or comparable to the BT1 phase except for Keith, who showed a slight increase.

SELF-INSTRUCTIONAL BEHAVIOUR

During the AT phase, all seven students learned to use the self-instructions, reaching 80% to 96% mastery by at least the third training session. The number of prompts required to keep the students adhering to the treatment procedure during the assessment sessions of this phase varied considerably across students but decreased across the phase for all students. In general division students required more prompts than multiplication students under all three conditions. Under the nc condition multiplication students required no prompts to keep them on the treatment procedure throughout the study, but division students required some prompts (mean = 2 - 3 prompts per session) only on the initial sessions of the AT phase. Prompts were required under both mc and ec conditions. Generally more prompts were required under mc than under ec in the earlier blocks for all students. Multiplication students required, on average, 4 - 6 prompts per session under mc and 2 - 5 prompts per session under ec. Division students required, on average, 7 - 8 prompts per session under mc and 2 - 4 prompts per session under ec. In the later blocks only 1 - 2 prompts per session were required for both mc and ec for all students. Virtually no prompts were required during the BT1 and BT2 phases, as students were very familiar with the procedure by that time.

Although the SI procedures were only taught in the context of ec and mc procedures, students spontaneously generalised the SI procedure under nc. Moreover, although self-instructions were taught only for ddm/ddd problems, all students generalised the three procedures to sdm/sdd and tdm/tdd problems except that, on occasions, division students did not generalise the mc procedure to sdd problems. This occurred only during the first two blocks of the AT phase. The procedure they used here was similar to that used in nc. Hence prompts were continued when necessary for the sdm/sdd and tdm/tdd problems only, for those spontaneously generalised procedures.

Student comments on the use of the procedures were recorded. Except for Mary, all initially expressed pleasure with mc, as they said it helped improve accuracy. However towards the later sessions, they expressed dislike at having to verbalise during assessment sessions and Keith expressed pleasure when told to use nc. Throughout the study, Mary expressed preference for the nc procedure even though she was aware that checking improved her accuracy. She also expressed strong dislike for verbalising while computing and became very resistant to it as the study progressed. Two of the seven students (Sue and Mary), reported having sore throats due to having to verbalise.

Discussion

Following SI training substantial improvements in accuracy on multiplication or division problems occurred with the two checking, and no-checking procedures for LD students who had mathematical deficits. Although no procedure was unequivocally superior across all students, accuracy was generally better under mc than under either ec or nc for all but Keith, for whom ec was generally superior. No consistent differences were found between ec and nc across students.

Under mc, each of the seven students showed rapid and substantial increases in accuracy on ddm/ddd problems (problems belonging to the trained cluster). These scores generally levelled off with some variability at high levels. This was accompanied by concomitant substantial decreases in error rate. Higher levels of accuracy were reached sooner with mc than with the other two procedures. Although the effectiveness of the three procedures was not unequivocally different, mc tended to be most effective for all students except Keith and Mary. Statistical analyses revealed that mc was significantly more effective than both nc and ec for three of the seven students. During BT1 accuracy levels were comparable to those reached in the AT phase for that treatment. Measurements taken 6 weeks later during the FU phase, revealed that generally, despite some decrease, accuracy levels were moderately high. However, they only returned to near 100% when the best treatment was reinstigated during BT2. This provided further support that mc was more effective than nc.

A number of generalisation effects were found. First, during nc, all students spontaneously utilised the SI procedure, although it was only taught in the context of mc or ec procedures. Second, although training was always provided employing ddm/ddd problems, all students spontaneously applied these procedures for untrained problem clusters, except initially with sdd problems under mc. Third, the effects of all three procedures generalised to untrained problems resulting in increased accuracy on all problem clusters. Again mc generally tended to be more effective than either ec or nc for the tdm/tdd problems. However, mc tended to be more effective than either ec or nc for sdm/sdd problems for only some students.

In the present study, accuracy, rate of completion and mean error rate were reported, as describing data in terms of percentage accuracy alone can be imprecise (White & Haring, 1976). Furthermore, as few problems were completed, small changes in behaviour would be greatly overestimated or underestimated if accuracy had been reported alone (Guilford, 1965). Evidence for this can be seen, for example, during block 1 of the AT phase, where Jane obtained 100% accuracy for tdd problems with ec, although she only completed one problem. Similarly under mc in block 5, she completed only two problems, one of which was correct, giving her 50% accuracy. Despite such individual variation however, some clear consistent effects were obtained. Error rates also helped in the analysis of the results. Rates (especially correct and error) are the only measures of performance that can provide information on proficiency and skill acquisition needs to be expressed in terms of proficiency as well as accuracy (Cooper et al., 1987). Also of all measures, rate is the most sensitive to the effects of environmental arrangements on performance because it reveals small increments of behaviour change (Cooper et al., 1987). It is possible to show percentage improvement even though incorrect responses are being maintained or even increased. For instance, although Keith's mean accuracy increased from 86.7% in the BT1 phase to 91.7% in the FU phase, his mean error rate remained constant for these phases.

Accuracy on the individual problem clusters was more variable than for the combined data (overall). This was probably largely due to the small number of each type of problem completed during each assessment period. Clearer patterns with less variability were obtained for the overall. As variability between procedures for ddm/ddd problems was sometimes large, data from overall accuracy will be discussed. Moreover, for both accuracy and rate, overall patterns generally reflected reasonably those obtained in the problem clusters.

While all students showed a substantial increase in accuracy following instruction, they also completed fewer problems, especially during ec and mc. It would be expected that if the students either overtly or covertly rehearsed an SI sequence while problem solving, the rate of responding would decrease. Also to accurately implement a problem solving strategy required more time than that taken by the nonsystematic behaviour observed in the students during baseline. This decrease in rate was consistent with previous findings that SI training decreases rate of responding (Burgio et al., 1980; M. B. Johnston et al., 1980; Whitman & Johnston, 1983). However, as suggested by Whitman & Johnston (1983), it may be that once self-instructions becomes automatised and/or covert, subsequent rate increases will occur. That there was a general increase in completion rate in the later phases of this study, even when self-instructions were overt, provide some evidence for this. However this can only be a tentative conclusion as it was not systematically assessed.

Besides indicating the potential of the SI error correction procedures for improving performance on relatively complex maths tasks, the present study extended the domain of such training to the educational needs of the learning disabled. The literature applying SI procedures to LD children is, to date, not plentiful. Furthermore, this is the first study to examine error correction which is student-initiated and which facilitates independent behaviour in the students. This investigation indicates that the procedures of self-instructions with error monitoring (especially when initiated by the student) hold promise, especially in light of current efforts in NZ to mainstream these children into normal classrooms. An influx of LD children into the regular classroom will require techniques that meet the needs of these children but do not overburden classroom teachers.

The instructional sequences used here dealt with fundamental maths skills, mastery of which is important educationally and socially for students with or without learning difficulties. Furthermore, the multiplication/division algorithms taught here were consistent with a sound modern maths orientation. For example, checking accuracy in addition and multiplication in the opposite direction illustrates the commutative property of addition and multiplication and checking subtraction problems by adding illustrates the inverse relation between subtraction and addition. It can, therefore, be argued that these sequences contributed to a foundation for future maths learning.

THE EFFECTIVENESS OF SELF-INSTRUCTION

That substantial improvement occurred with or without error checking procedures shows that SI training was a crucial part of all procedures. This provides further support for previous findings (Albion & Salzberg, 1982; M. B. Johnston et al., 1980; Leon & Pepe, 1983; Whitman & Johnston, 1983; Van Luit & Van der Aalsvoort, 1985) that self-instruction is an effective training procedure.

There were several components in the SI training that may have contributed to its effectiveness. The present study included three major components of the SI package namely modelling, verbalisation, and strategy (in the form of self-instructions). Each of these components has been shown to be effective in previous research. However, in the present study, the relative efficacy of each component in the SI package cannot be unequivocally determined as this was not systematically investigated.

The present findings show that modelling can be an effective instructional intervention that enhances accurate and efficient learning among LD students. The SI procedure involved essentially a demonstration and imitation of a strategy in the form of key-instructions (verbal and motor) intervention. The demonstration component provided systematic presentation of the targeted skills supporting previous research findings of its crucial role (Blankenship, 1978, Blankenship & Baumgartner, 1982; Rivera & Smith, 1988; Smith & Lovitt, 1976). The imitation component may also have been important since it may allow assessment of whether correct learning had occurred (Rivera & Smith, 1988).

The use of a strategy may be a potent component of the SI package. R. N. Roberts et al. (1987), in demonstrating that SI was no more effective than strategy training involving a straight operant procedure of reinforcing arithmetic accuracy, suggested that the utility of SI may lie in providing step-by-step clear instructions that leads to the solution of the problem rather than modifying cognitive processes. In the present study, although all three experimental procedures differed, they had a strategy in common. A strategy entails a series of steps that leads to the solution of a problem. Some support for the effectiveness of strategy comes from the performance of the division students. They were unable to solve problems with a zero in the quotient when using the SI sequences until they were provided with a further strategy in the first session of the third block. Immediately after that their accuracy improved under all three conditions. The strategy, by providing sufficient explicit structure may have enhanced stimulus control over incorrect responses. That a strategy can be highly effective is supported by previous research (e.g., Lloyd et al., 1981; Montague & Bos, 1986; Rivera & Smith, 1988).

The effectiveness of self-instruction may also be explained, in part, by looking at how the prior verbalisation influences written behaviour. It has been suggested that overt verbalisation may alter the performance by adding another stimulus dimension to the problem (Lovitt & Curtiss, 1968), by making the performance more accurate and deliberate (Lovitt & Curtiss, 1968), or by directing the subject's attention to important task features (Schunk, 1982). Research has also indicated that self-verbalisation can produce discriminative stimuli that directly increase the probability of a correct non-verbal response (e.g., Hartig & Kanfer, 1973).

Verbalisation may also affect performance in part by focusing attention on the task, that is, it may exercise control over precurrent behaviour. Some support for this is provided by studies which showed that overt (audible) self-instructions were more effective than covert (silent) self-instructions (Hartig & Kanfer, 1973; Meichenbaum & Goodman, 1969). Although no systematic attempt was made in the present study to measure students' on-task behaviour, Sue, and to a smaller extent Joe, were sometimes observed to be off-task. For instance, on occasions they day dreamt or forgot to add the partial products of their multiplication problems, when not verbalising the strategy (during the BL phase). These behaviours decreased, during all other phases when they were required to verbalise. However, to clearly show an "attention" function it would be necessary to demonstrate more directly that overt verbalisation improves attention to the task and not simply task performance. In the present study, results from the VB phase suggest that verbalisation had no effect on accuracy before instruction. It could only be effective when students had the prerequisite skills to solve the problem which may explain why verbalisation has produced discrepant results in the literature (Cullinan et al., 1978; Lovitt & Curtiss, 1968; Parsons, 1972). If children did possess such skills, but were frequently off task, increasing attention alone might be sufficient to improve academic performance.

A number of other factors may also have influenced the effectiveness of the SI training procedures in this study. First, during training, correspondence between the words verbalised by the student and the motor behaviour was explicitly reinforced. A considerable body of research (R. A. Baer et al., 1987; Israel, 1973, 1978) has indicated that strengthening of the correspondence between verbal and motor behaviour enhances performance.

Second, the present study also incorporated such operant techniques as prompting and reinforcement. The systematic utilisation of these procedures to teach self-instructions may be an important factor in SI training packages since their effectiveness is documented in previous research (Grimm et al., 1973; R. N. Roberts et al., 1987; Whitman & Johnston, 1983). Moreover, some studies (Friedling & O'Leary, 1979; Varni & Henker, 1979) found SI training to be ineffective until a token system was subsequently introduced.

Third, the use of self-instructions enables a subject to work at tasks in a more systematic manner (Wilder, Draper and Donnelly, 1984) and ensures consistency in teaching. For example, there are several ways of solving a long multiplication/division problem, and at school, children are often inadvertently taught more than one method. In this study, however, the children were presented with only one approach to working the algorithms, and this approach was carefully specified. Thus consistency and the use of a systematic approach may have contributed to the treatment effects.

Fourth, the specificity of self-instructions may also contribute to the effectiveness of SI. The self-instructions used by Meichenbaum and Goodman (1971) were not sequences of arithmetic operations but involved recalling the general approach to the problem, going slowly and carefully, correcting errors, reinforcing participation as well as successful work. These instructions ("go slow" and "be careful") were less explicit than those used in the present study and may not be useful in some cases. Such self-instructions are not likely to lead to success if the student does not have the necessary skills, no matter how well they are followed. This is especially true of self-instructions of the form, "If I try harder, I'll get it right." For a given student, there are always problems for which no amount of harder trying will accomplish a correct answer. Self-instructions such as "go slowly", "be careful" and "try harder", function more to mediate contingencies of reinforcement, punishment and extinction (D. M. Baer, 1982). This means that their function would be motivational, rather than to provide task-related discriminative stimuli.

Although the SI program used in this study had several elements in common with most previous SI studies, it differed in several significant respects. First, the present study used the self-evaluation component as one of the principal components unlike previous studies. It systematically varied this component by comparing two checking procedures and a no-checking procedure. Second, the self-evaluation component in most previous research usually included general statements (e.g., "What is the plan and am I following it?") but in the present study it was a means of specific error monitoring and correction. Error-monitoring as used by M. B. Johnston et al. (1980), and Van Luit and Van der Aalsvoort (1985) was separable from the strategy, because they required it to occur after the strategy had been fully implemented. This was done with ec in the present study. In nc, formal error-monitoring was not allowed to be implemented but in mc, error-monitoring was ongoing, occurring throughout the strategy. Third, unlike previous studies, the procedure required verbalisation to be overt during all assessment periods (except during the BL phase) for purposes of validation. Fourth, teacher prompts were not faded whereas in other studies they were always faded.

The self-monitoring literature has sometimes used the terms "self-evaluation" and "self-monitoring" synonymously. The checking (error-monitoring) procedures used in this study are a form of self-monitoring. The procedures are also self-evaluative in that they are assessing the accuracy of the students' accuracy. However, the self-monitoring procedures used here differed from previous self-monitoring research in two important aspects. First, they did not monitor on-task behaviour. Previous studies (Hallahan et al., 1979; Heins, 1980; Lloyd et al., 1980) have attempted to increase on-task behaviour of students with attentional problems in order to improve academic productivity, but have produced minimal gains in accuracy. Students who had attentional deficits were not used in the present study as attention was needed to learn the self-instructions. Second, monitoring of academic variables in previous research, involved students comparing answers to solved problems with the answers provided by the experimenter and recording whether each problem was solved correctly or incorrectly (Rooney et al., 1985). In the present study, monitoring of the academic variable was more active, in that the student actively monitored his/her performance by checking and correcting errors on his/her own.

In some self-instruction research (Leon & Pepe, 1983, Varni & Henker, 1979) neither the specific target maths tasks nor the SI sequences were adequately presented. It is difficult to determine whether performance gains were due to the SI sequences used or to the type of maths tasks involved. Clearly, if SI is to be adequately evaluated and developed, the specifics of the procedure must be articulated (M. B. Johnston et al., 1980). Hence the self-instructions taught in this study were specified fully.

The role of self-instruction is questionable in situations where it is not specifically prompted or carried out overtly (e.g., M. B. Johnston et al., 1980). According to S. G. O'Leary and Dubey (1979) telling a child to self-instruct does not always guarantee adherence. The subjects in the present study were required to verbalise the SI procedures throughout the experiment. Furthermore, experimenter prompts were given to keep the students adhering to the treatment procedure to ensure that the changes in performance were a function of the particular procedure. The experimenter's instructions to use self-verbalisation resulted in high rates of appropriate self-instruction use, as evidenced by the small number of prompts required to keep them following the procedure and the substantial concomitant increases in correct responding. The students' mastery of the procedures continued to improve across the phases since the number of prompts required to keep them on task decreased. The findings of the present study allow more conclusive statements about the effects of self-instructions in the classroom than previous SI studies that have not assessed acquisition of SI skills as a function of training (Leon & Pepe, 1983), or that have not measured SI use by making it overt (e.g., M. B. Johnston et al., 1980).

Research (R. N. Roberts, 1979, R. N. Roberts et al., 1987) has shown that once an academic task was learned, no inhibitory effect occurred due to using overt self-verbalisation. Although this was not systematically investigated in the present study, the students frequently reported finding verbalising aversive especially as the study proceeded. There may have been several reasons for this. First, because the tasks used and the assessment periods were short in previous studies, verbalising may never have become aversive. Second, most previous studies did not require students to verbalise throughout the whole assessment period. Third, the subjects used in most of the previous research were young. Older students may be less willing to verbalise repetitive instructions.

The literature on self-instruction seems particularly susceptible to being analysed in cognitive terms (Brigham, 1980). Private events need to be analysed thoroughly as physical, behavioural events (J. M. Johnston & Pennypacker, 1980; Moore, 1980, 1984; Skinner, 1945, 1953, 1957, 1974). In this study self-instructions were assumed to be capable of playing a role in behaviour as mediators (Skinner, 1974, p. 325), or parts of causal chains of behaviour (D. M. Baer, 1982; Moore, 1984). Some of these events were able to be investigated in the present study by requiring the students to verbalise the self-instructions.

The SI procedure used in the present study comprised a problem-solving strategy. According to Skinner (1953), problem solving is not concerned simply with emitting a solution but rather with the techniques of finding the solution. Thus solving a problem begins with an external stimulus having a discriminative function, which gives rise to the generation of further discriminative stimuli in the form of observable and nonobservable stimulus-response-stimulus interactions, and ends with an external response and a reinforcement contingency.

Grimm et al. (1973) has suggested three reasons why a student may fail to emit a complete problem-solving chain. First, it may be due to a limited behaviour repertoire or lack of appropriate conditioning history as when a student is unable to count because she/he cannot tact written numbers. Second, conditioned reinforcers, possibly including the terminal reinforcer, in the chain may be weak or nonfunctional. For example, a student, for whom social approval is not a reinforcer, is praised for completing an arithmetic problem. Here the chain will not be strengthened or maintained. Third, weak or absent discriminative control of appropriate responding by stimuli in the chain may result in the chain not developing. For instance when a given stimulus controls a competing response like doodling instead of writing. These conditions may combine in various ways to prevent progress in academic situations which require lengthy response chains. The study presented here dealt with the third type of deficit – weak or absent discriminative control by stimuli in the chain.

Smith and Lovitt (1976), have shown that if a chain weakened, by weak antecedent stimulus control (due to lack of preskills), the problem-solving response will in all probability not improve performance due to a very low reinforcement rate. Hence, the components of the chain require strengthening. If these components are covert, making them overt would enable them to be monitored. This would then allow appropriate antecedent variables to be applied and correct responding to be reinforced. After the chain is sufficiently strengthened by this procedure, these special contingencies should be gradually removed so that the entire chain is supported by the reinforcement provided by obtaining the correct solution. In the present study the chain was strengthened by using a checking procedure. Such a "correctional" procedure should assist a student to deal more effectively with subsequent arithmetic problem-solving tasks.

The self-monitoring literature often seems to imply that simply monitoring behaviour causes it to be altered in a positive way. However, monitoring alone will often not be sufficient (Brigham, 1980). For instance, in monitoring for errors, students may not detect the errors, or alternatively may detect, but not correct, errors. Whether these errors are corrected after being detected will depend on the student's prior experience and training with regard to those specific errors (Brigham, 1980). In the present study it was likely that students would correct errors when they were detected as during training the experimenter modelled the correction of errors detected during checking. Furthermore, such behaviour was reinforced verbally when it occurred. It seems likely that detection of errors was reinforcing because it enabled correction of these errors. This detection and correction of errors in the links would increase the probability of getting the correct final answer, which is arguably the terminal reinforcer. It is reasonable to assume that the three treatment procedures could have produced different rates of error detection and hence correction. This would have resulted in different reinforcement rates in the links under the different treatment procedures. The use of error correction procedures particularly mc should have increased the strengthening of links and hence the probability of terminal reinforcement. This may explain why mc generally resulted in the highest accuracy on most occasions. These error-monitoring procedures might be expected to be especially valuable for LD children who have been described as learned-helpless (Gresham, 1988), since the relationship between actions and outcomes are highlighted (Ryan et al., 1986).

DIFFERENTIAL EFFECTIVENESS OF THE PROCEDURES

The overall accuracy scores generally showed much clearer experimental control by mc, than the accuracy scores for ddm/ddd problems. Clear experimental control by mc was shown for four students. Statistical analyses revealed that mc was significantly more effective than either ec or nc for these students and one other student. Included were the three students for whom mc was significantly most effective for ddm/ddd problems and two others. For one student, Mary, no significant results were found. However, mc produced the highest overall accuracy scores on most blocks throughout the AT phase. For all these six students mean overall error rate was lowest under mc. However, Mary's results are suspect as she was resistant and uninterested throughout and she did not complete the study.

That mc produced higher accuracy scores than ec and nc was as expected. The greater effectiveness of mc, compared to nc, is further substantiated by the overall accuracy levels being higher and overall mean error rates lower during the BT1 and BT2 phases compared to during the FU phase. This was consistent with previous research that showed higher accuracy under error correction than under no correction (Carnine, 1977, cited in Rosenberg, 1986). However, no clear conclusions can be made about ec in relation to nc. For Keith, ec was most effective in most blocks for both ddm problems and overall.

However, the degree of difference between mc and ec was not large, and on ddm problems the mean accuracy was actually higher under mc than under ec. Thus in general it can be concluded that mc tended to be the most effective procedure for all students. All three procedures differed only in whether or not checking occurred and, if it did, the method of checking. This aspect of the SI procedure most likely accounted for the differential effect of the three procedures.

Most links in the problem solving chain consisted of checking combination facts. An incorrect combination could occur involving any of the four basic operations (Grossnickle, 1936). Though it was not the focus of this study, error analysis revealed that during baseline most students made both algorithm errors and combination errors. A faulty procedure (Grossnickle, 1936) or an algorithm error involves the application of incorrect rules or strategies in solving the problem (Drucker, McBride, & Wilbur, 1987). These errors do not include basic fact errors. After instruction all students usually made more combination errors than algorithm errors. This is consistent with previous findings (Grossnickle, 1939; Reisman, 1982) which showed that the commonest errors are combination ones. As most links involve checking combinations, it is uncertain whether mc would be most effective for detecting other types of errors.

The tendency for greater effectiveness of mc over ec and nc may possibly be explained as follows. Under mc, frequent assessment and the immediacy of checking may have permitted early and easy identification of errors, which then allowed for correction of these errors. Under mc, checking each link (by using the inverse operation, or the same operation but in the reverse direction) could have strengthened each link by increasing the probability of appropriate stimulus control in the next link, thus maximising the probability of getting the correct final answer. However, all links needed to be correct for the right answer to be obtained. Checking each link was likely to be reinforcing only when errors were made and then detected and corrected. To be able to detect combination errors, students required a reasonably high basic facts knowledge. It was observed that, especially during the acquisition stage, some multiplication students used an adding procedure for deriving multiplication facts that they were unsure of, especially when asked to multi-check. Later they dispensed with such a procedure, possibly because these newly acquired multiplication responses were strengthened. Mc, in providing frequent assessment of learner progress within each problem, may have increased the rate of error detection and correction which is reinforcing, and further strengthened these responses. That frequent assessment of learner progess is reinforcing and strengthens newly acquired responses has been previously suggested by Baine (1982). During ec, it may have been more difficult to detect errors embedded in a long chain because of too many stimuli. This may have resulted in lower rates and longer delays of reinforcement. Furthermore, appropriate stimulus control may be less probable under ec as checking was not done using the inverse operation or the reverse direction. For nc, even fewer errors are likely to be detected as no explicit checking was carried out. However, no consistent superiority was found between ec and nc across students.

It is difficult to directly control thought processes. For instance, Grossnickle (1938), in studying the effectiveness of checking subtraction by addition, found that students often pretended to check and so no difference in accuracy was found between those who checked and those who did not. However, Grossnickle did not require his students to check overtly. Cognitive research has shown that it is difficult to verbalise while thinking about something else. However, this possibility cannot necessarily be eradicated totally. In the present study, when using mc, there may have been occasions when verbalisation merely involved echoics and not appropriate intraverbals or accurate tacting (Skinner, 1957). It was often suspected that a student emitting say, 6x3=18 repeated 3x6=18 overtly but with the second 18 being an echoic rather than part of the intraverbal 3x6=18. There was no way of measuring this. However, intraverbals would be more likely when the student becomes proficient in the basic facts. In fact, the students made few errors with this type of basic fact in the later problems. Evidence for an echoic response in the case of addition would be, say (4+7+9=21) where 9+7+4=21 is said as a repetition. Twenty-one is probably an echoic response, as the student with the necessary preskills in addition would have detected the error if he/she had been under intraverbal control when recalculating. A procedural change that could have helped for checking addition in the reverse direction would be to require the students to calculate and verbalise partial answers, so reducing the probability of echoics. For instance, the sum 4+7+9=20 would be solved by 4+7=11; +9=20 and then checked in the reverse direction by requiring verbalisation at 9+7=16; +4=20. At times wrong verbalisation occurred during mc where students verbalised 0x7=0 and wrote 0 and then when multi-checking verbalised 7x0=7 without noting what he/she had written down.

An attempt was made to ensure appropriate tacting and intraverbals with the division students by requiring them to use frequently the inverse operation to check (e.g., checking subtraction by addition). This was not done for the multiplication students as they did not have sufficient preskills to do the more difficult inverse operation (division). Hence there were more opportunities for them to make echoic responses. However, division students also were not taught to use the inverse operation when checking multiplication as they were not able to solve problems with zeros in the quotient. Though division students used the multiplication operation in the reverse direction to obtain and check the largest multiple of the divisor, they rewrote and remultiplied each term of the quotient and the divisor while checking, and compared it to the initial product they obtained (see division script in

Appendix 2). This procedure also may have helped prevent echoic responses and increase tacting.

Keith's results can also be considered in the light of the general explanation being developed. The greater effectiveness of ec for Keith may be explained as follows. Keith had the highest mastery of prerequisite skills and facts, and therefore may not have made, and hence detected, as many errors with mc as the other multiplication students initially. Although he made combination errors, his main errors were algorithm ones (i.e., zero place-holder errors). After instruction, he made no algorithm errors. Since he did not make as many combination errors as the other students, he may have received less differential reinforcement under mc. This hypothesis is consistent with the results of the division students who, although possessed higher multiplication fact knowledge, had lower computational skills in subtraction.

If checking under mc increased the rate and probability of detecting errors and hence reinforcement, mc should be especially effective for students with reasonable preskills but who are careless or impulsive and who make many combination errors. However, students with low mastery levels of skills and facts may not be able to detect errors.

If it is true that error detection, which allows error correction, is reinforcing, then ec would be expected to result in higher accuracy than nc. However in the present findings nc and ec produced no consistent differences in accuracy across students with ddm/ddd problems. For overall accuracy, ec produced higher scores than nc on most blocks for four students. The reverse was true for two students and one showed no consistent difference between nc and ec. However there are some factors that may have affected students differentially. First, appropriate stimulus control may have been more difficult to establish with ec due to the number of stimuli created by 'carry-over' numbers during the initial solution. This was observed with some students, particularly multiplication students. They had great difficulty in deciphering which 'carry-over numbers' belonged to the particular multiplication fact they were checking. This could have resulted in incorrect corrections for some. Thus these subtasks had to be broken down even further to reduce difficulty in establishing correct stimulus control. This was done, for example, by showing them exactly where to place the 'carry-over' numbers of each partial product to avoid confusion later whilst checking. Yet, despite this, some of these students still made mistakes. For such students, it is probably advisable to check by rewriting the whole problem again. Second, some students may have not checked fully, that is, they may not have checked the answer with that previously written. For instance on some occasions, Vera, when initially solving the problem did a link incorrectly (e.g., 7x3=21, +4=26) but when checking verbalised the correct answer (7x3=21, +4=25), and yet did not correct it. This was because she did not detect the error as she did not check with what was previously written. Third, some may have emitted textual responses (Skinner, 1957). For instance, when calculating a problem, Joe would verbalise 'seven times three is twenty-one, plus four equals twenty-five' (7x3=21, +4=25) and would write 5 down and carry the 2 over to the next column. However, when he end-checked he verbalised it as 7x3=25. Third, students may have varied in the amount they checked combination facts for each partial product/dividend. Though all students carried out the ec procedure as taught and checked at least half of the combination facts for each partial product/dividend, some consistently checked every link (though this was not required). Moreover, there were instances where some would verbalise a product (say 7x4=28) without attempting to add on the carry-over from the previous multiplication (say +3) though they wrote down the addition (i.e 31), and then proceeded to check the next part of the problem. On the other hand, the addition of the carry-over number may have been done subvocally. Definite conclusions are difficult as all students, especially those doing multiplication, varied somewhat in the way they end-checked. It seems likely that ec would be only as effective as, or even less effective than, nc, if incorrect corrections were made to the problems. Thus, the results need to be interpreted with caution.

As students became more proficient under their best procedure, accuracy, for ddm/ddd problems and overall, stabilised at high levels. Generally the accuracy levels of the other two procedures tended to stabilise close to, although lower than, the best procedure levels during the proficiency stage. Thus it would seem that the three procedures affected accuracy differentially mostly during the acquisition stages of skill learning and less so during the proficiency stages, with ec being no more effective than nc during the proficiency stage. Further evidence that ec may be no more effective than nc during proficiency was provided by Keith who showed no appreciable change in accuracy during the BT1, FU, and BT2 phases. This has implications for teachers who always require students to end-check as it may no longer be useful when students become proficient.

This differential effect between procedures during the different stages of skill acquisition, may be explained as follows. During skill acquisition, students were liable to make many mistakes. More of these errors were likely to be detected under mc because of the high frequency of checking. Hence, despite the low completion rate, reinforcement would likely be greatest under mc. However, as they became more proficient, fewer errors would probably be detected and hence corrected. Thus, despite the high frequency of checking, the differential rate of reinforcement under mc during the proficiency stage would be reduced. Mc would thus not be expected to continue to produce as large a differential effect during the later stages of skill acquisition. Yet, despite this, mc still produced the highest scores. It is unclear why the differential effect between the procedures for Keith did not decrease appreciably towards the proficiency stage or why accuracy levels did not level as quickly as for most of the others.

While some possible reinforcing properties of the procedures have been considered, some aversive properties might also be present. There is little doubt that greater effort was required in using checking procedures, in particular mc. Requiring students who have become proficient to check as frequently as before may have limited the effectiveness of mc. During the acquisition stage the reinforcement obtained for detecting many errors may have outweighed the punishing effects of low completion rate. During the proficiency stage, however, nc (and ec to a slighter extent) may have been more reinforcing as more problems were being completed with high accuracy levels. This may explain students' reported preference of the nc procedure despite its lower accuracy levels to mc. In spite of this, mc was still better, although not as differentially effective as during acquisition. Moreover, if errors recur during the proficiency stage, mc would still be more effective than nc or ec in detecting these errors.

Most students usually levelled off at levels lower than 100% and reported being happy with 70%-80% accuracy. This was possibly due to two factors. First, a history of repeated failure to gain complete mastery may have contributed to a lack of motivation and hence extrinsic motivation would be needed to reinforce reaching 100%. Second, low expectations of teachers throughout school life (where 70% was considered good), may not have made 100% mastery more highly reinforcing.

There was some evidence that mc may have been more useful for slower students with attention problems. For instance, the differential effect between procedures was greatest with Sue, who was labelled by the school as a slow learner. In addition, she was the most distractible student (when compared to the others) as shown by her ratings on the Conners (1973) scale. Evidence for her being a slow learner can also be seen in that during the AT phase, her accuracy scores did not plateau as quickly as did those for most of the others. Mc showed the clearest effect for her in both the ddm and overall problems.

If mc's greater reinforcing properties are due to the detection and correction of more errors, it follows that the length and complexity of the chain may be important. In longer and more complex chains, mc may be more effective since there would be a greater likelihood of errors being made and detected. Evidence for the general effectiveness of mc with long and complex chains was provided by both multiplication and division students on the multidigit multiplier/divisor problems. However, the results of the division students and one multiplication student which showed mc to be most effective with the simpler problems, does not support this hypothesis. Three reasons are suggested. First, the results are confounded to some extent in the initial sessions due to the lack of spontaneous generalisation of mc for these problems. However, there is a possibility that students may have done mc subvocally during these sessions. Second, half the sdd problems used in the worksheets were quite difficult as they had one or more zeros in the quotient and thus mc was effective with such complex problems. In fact errors, made on these problems were nearly always due to zero errors. Research has shown that such problems are considered one of the most difficult (Brueckner & Melbye, 1940). Third, Sue was a slow learner and mc may be effective even with simpler tasks for such students.

The tendency for mc to produce the greatest accuracy occurred irrespective of the tasks (multiplication or division) and despite any qualitative and quantitative differences between them. However, the effectiveness of mc is clearer for division where higher levels of accuracy were maintained than for multiplication. In practically all phases division students produced higher scores than multiplication students, usually reaching 100% for both trained and untrained problem clusters on several occasions with mc. Three reasons are suggested for this. First, task complexity was a factor as division involved longer and more complex chains than multiplication (consisting of a considerable variety of skills in combination, varying widely in difficulty). Secondly, the division students had a higher level of preskill knowledge than the multiplication students. Thirdly, procedural differences in the checking between these two groups could have had an effect. The division checking procedures may have been more beneficial since checking, by using the inverse operation, increased the likelihood of tacts occurring rather than simply echoics. If the inverse operation had been used by multiplication students, clearer results may have emerged.

Multiplication students completed the largest number of problems during nc and the least during mc. This is not surprising as more time was needed for each problem during mc than nc or ec because of the repetition of each link. This however, is not the case for division students. Although both division students completed the largest number of problems under nc, the lowest number completed was usually under ec. The reason for this is unclear. Moreover, it is noted that division students completed a smaller number of problems under each condition than multiplication students. One possible reason is that the division problems involved longer and more complex chains than the multiplication problems. In spite of the low completion rates, higher levels of accuracy were attained. Though it may not seem to be cost effective in terms of the completion rate, it is suggested that reinforcement of accuracy is more important than rate, at least initially, and especially for students who have often experienced failure.

GENERALISATION

The problems employed during the training period each day were similar but not identical to the ddm/ddd problems presented during assessment. Moreover, problems which belonged to this cluster but which varied in the number of digits in the multiplicand/dividend were also presented during assessment. Despite the very small number of problems used during training and the very short training period, students' accuracy increased for these untaught within-cluster problems. This supports previous research (Blankenship, 1978; Smith & Lovitt, 1975) which showed that students generalise learning to similar, uninstructed problems. Furthermore, during maintenance checks in the FU phase, given six weeks after the BT1 phase, accuracy levels were above baseline levels for both groups of students.

Increases in accuracy also occurred with sdm/sdd and tdm/tdd problems, problems not specifically trained in this study. For all students this generalisation occurred spontaneously, irrespective of the procedure used, and even though, nearly all students scored 0% on tdm/tdd problems during the BL and VB phases. Thus when a problem of moderate complexity (ddm/ddd) was targeted for instruction, learning generalised to both easier (sdm/sdd) and more difficult ones (tdm/tdd) even without direct instruction with these problem types.

Previous research had shown that generalisation to dissimilar problem types was not successful if instructed problems were easier (Smith & Fleming, 1977, cited in Rivera & Smith, 1987; Smith & Lovitt, 1975). Generalisation to similar and dissimilar types of problems was enhanced by instructing the most difficult problem type (Rivera et al., 1984, cited in Rivera & Smith, 1987; Smith, 1978, cited in Rivera & Smith, 1987). The present study showed that targeting a problem of medium complexity can be sufficient to produce generalisation to problem types of both greater and less complexity. How the complexity of instructed problems affects generalisation needs further analysis but this would be useful if instructional time can be saved (Rivera & Smith, 1987). One possible reason for the discrepant results could be that different forms of instruction were used. Previous research used D&PM as their form of instruction, not verbalised self-instructions as used in the present study.

Although SI training was given with the ec and mc procedures, all students spontaneously generalised the procedure under nc to all problems whether of the trained or untrained clusters. Generalisation (under nc) was probably greater from ec as nc was similar to the ec procedure but without the checking part at the end. Multiplication students generalised all three procedures to the sdm and tdm problems. As generalisation was spontaneous, the prompting procedure was also used for these problems when required. The division students, however, spontaneously generalised all three procedures to the tdd problems but only ec and nc to the sdd problems. Hence prompts were continued when necessary for the sdd and tdd problems only, for those spontaneously generalised procedures.

Generalisation of mc to sdd problems occurred but only after the first block of the AT phase.

Taken together, these data show that changes in accuracy generalised across task stimuli (variations in number of digits in the multiplicand/dividend but of the same class of problems), task type (problems belonging to different clusters), and time. This supports the conclusions of Stokes and Baer (1977), about the potential of verbal stimulus control procedures for programming generalisation. It can be argued that these generalised changes occurred because the self-instructions implemented in this study were designed to provide the students with a general problem solving strategy applicable to a wide variety of multiplication/division tasks. That is, effectively, generalisation was programmed. For instance, during training with ddm problems multiplication students were taught that when calculating the second partial product, only one zero needs to be placed at the end when multiplying by the tens digit. Immediately after instruction students generalised this rule to tdm problems by placing two zeros at the end when multiplying by the hundreds digit to obtain the third partial product, when initially they had made such errors during the BL and VB phases.

Also needing analysis is identification of those factors which impede generalisation and those which foster it (Rivera & Smith, 1987). Blankenship and Baumgartner (1982) suggested that different types of interventions may be necessary to enhance generalisation depending on student ability and motivational level. Although generalisation can occur with LD students after they have received D&PM interventions, definitive statements about enhancing generalisation cannot be made as they do not do so consistently (Rivera & Smith, 1987). Cullinan, Lloyd and Epstein (1981) recommended that when alternative attack strategies can be used for a particular task class, the instructional designer should select those that include preskills that overlap with preskills used in other task classes. This suggests that the ability to generalise may be a function of how many skills learnt with the trained problems can be used with the untrained problems. For instance division students continued making errors in problems having one or more zeros in the quotient until specifically trained on such problems. After instruction, error rate on such problems reduced to zero immediately across all problem clusters. Generalisation with such problems did not occur until this particular skill was taught.

The degree to which each treatment effect generalised, varied among the problem clusters. For instance, students generally (with a few exceptions) scored higher levels of accuracy for sdm/sdd problems than with tdm/tdd problems for all three procedures during the AT, BT1, and BT2 phases. There are various explanations for this variation. First, it could have been caused by the nature of the generalisation problems themselves which varied in complexity. Furthermore, there were limited computational variations possible for sdm/sdd problems than tdm/tdd problems. This additional factor may have accounted for the enhanced generalisation for sdm/sdd problems. Second, it could be that variability is accentuated by reduced sample size. Third, it might be that the ability to generalise may be related to the degree of task mastery on instructed type problems (Blankenship & Baumgartner, 1982). Furthermore, it is also difficult to assess the extent of the generalisation effect on sdm/sdd problems because of the moderately high levels of proficiency of some students on these problems during baseline.

Self-instructions can be effective self-controlling procedures only if students actually implement them to influence the behaviours being changed (S. G. O'Leary & Dubey, 1979). Providing reinforcement assists adherence to their self-instructions (S. G. O'Leary & Dubey, 1979). In the present study, use of self-instructions as a self-control procedure was not reinforced since the main focus was to determine the relative effectiveness of the three procedures. This might explain why students did not use the checking procedures when given a choice during the FU phase even though they had reported that it was good to check as it improved their accuracy. Moreover, it is possible that towards the later stages when students became more proficient, checking was insufficiently reinforced because few errors were being made and hence detected.

LD SUBJECTS

NZ studies pertaining to learning disabilities have employed definitions involving the achievement-ability discrepancy factor, and restricted themselves to subject areas that may be assessed with available NZ standardised tests (Chapman, 1985). The present study used a similar approach. The students used were all classified as LD.

However, the whole question of such classification is ambiguous. There is considerable disagreement as to how to define or assess learning disability. A number of investigators have made efforts to categorise LD children into more homogeneous subgroups based on specific academic patterns (e.g., Barkley, 1981) Although these subclassification schemes represent attempts to reduce the heterogeneity covered by the term "learning disabilities" they have not yet received strong empirical support (Treiber & Lahey, 1983).

A strong case for the communality of effective instruction for the ED, EMR and LD populations can be found within the behaviourist tradition. The advocate of behaviour modification is less concerned with the diagnostic category in to which a child has been placed, and more with behaviours exhibited, as the basis on which the teaching strategy should hinge (Hallahan & Kauffman, 1976).

The rationale for adopting such a broad view of learning disabilities is supported by evidence from the literature. First, in current public policy, operational criteria for distinguishing discrete categories of mildly handicapped students are sufficiently vague and arbitrary as to diminish their importance for the purposes of reviewing the instructional and programmatic literature (Ysseldyke, Algozzine, Shinn, & McGue, 1982). Second, the actual definition of learning disability employed by different investigators varies and many of the children who were not labelled as LD in behavioural studies would probably meet the federal criteria for learning disability (Federal Register, 1977, cited in Wong, 1986). Third, even if subgroups of poor learners can be rigorously identified, evidence from applied research seems to support the use of highly similar instructional principles and methods for correction (Hallahan & Kauffman, 1976; Reschly, 1987; Morsink, Thomas, & Smith-Davis, 1987). Moreover, it has been argued that theoretical discreteness in categorical labels need not imply the need for correspondingly discrete instructional interventions. Hallahan & Kauffman (1977) drew attention to the practically-significant overlap in instructionally relevant behavioural characteristics that are attributed to different categories of mildly handicapped students. Therefore, instructional differences are assumed to be most usefully characterised as differences in how intensively or extensively various methods must be applied to achieve reliable academic gains (Gerber, 1987). Fourth, specialists in the field of LD, EMR, and ED have been recommending methods that matched children's behaviours rather than their categories (Morsink et al., 1987).

In fact some (e.g., Kavale & Forness, 1985) have argued for a "paradigm shift" that would require a multidefinitional perspective (Chapman & Wilkinson, 1988) in the learning disabilities field. This would "involve a broadening of the concept of learning disabilities to embrace conceptions of learning failure rather than focus on traditional notions based on hypothetical intrinsic processes" (Chapman & Wilkinson, 1988, p. 12). In a similar vein, Farrald and Schamber (1973) proposed that LD children "require extensive remodelling and reconstruction of *teaching* interventions for efficient teaching" (p. 20). Others (Hallahan & Kauffman, 1977; Ysseldyke, Algozzine, & Thurlow, 1983), in asserting the uselessness of the concept of learning disability, have suggested that the classification stage be skipped because the classification of learning disability has little or no instructional validity.

LIMITATIONS AND FUTURE DIRECTIONS

There were several problems associated with the present study. One practical one was with the overt verbalisation of self-instructions which was required for validation purposes.

Most students found verbalising the procedures aversive with some students reporting sore throats on some occasions. This is not surprising as they had to verbalise the procedure for 20 minutes, often twice a day. In a teaching situation, only a short time would usually be required to ensure correct use of the procedure. To alleviate the aversiveness of verbalising during the assessment period, verbalisation was reduced to 12 minutes, and pointing only was required for the last 8 minutes. Students reported being relatively happy with this change. However, no reliability checks could be made when students were pointing as a video recording was not available. Towards the end of the study, there was also difficulty in motivating the students to continue as they reported getting bored. Despite adverse verbal comments, mc still produced the highest scores on most blocks. Future research should look at ways of overcoming the aversiveness of verbalisation.

One important limitation is the unreliability of verbalisation as an indicator of the covert processes. There were instances, though few, of wrong verbalisation, especially inappropriate tacting. Attempts could be made in future research to ensure that students' verbalisation correspond with their written behaviour.

Another limitation of this study was due to the ec procedure not being implemented identically by all students. Although there was some structure in the way ec was taught, it appears that even more structure should have been imposed. Certainly, teachers need to make an effort to teach students how to end-check. Often teachers recommend checking the answer by doing the inverse operation. This can provide a reliable and quick check but may not be feasible for students who do not have the preskills for using the inverse operation. When teachers do suggest end-checking the whole problem, they do not necessarily show how this is to be done. The present study demonstrated that subtasks need to be broken down even further during ec, to reduce difficulty in establishing correct stimulus control. Clearly the ec procedure could be more highly structured in future research especially as it is a checking procedure widely encouraged in the classroom.

The small number of problems in each cluster completed by students may have increased variability in performance in the clusters making comparison of the effect of the various procedures difficult, especially in the AT phase. These problems could have been reduced by the use of large and equal number of problems from each problem cluster. This would have overcome the confounding variable of time (which affected both error and correct rate). In the present study it was not practical to give an equal number of problems for each cluster (e.g., by using different worksheets for each problem type). This was because slower students would have taken too much time to complete the problems which would have infringed too much on the school's time and they might have become bored and fatigued. The students were not allowed out of normal classes for long periods which also accounted for the small number of sessions given in the BT1, FU and BT2 phases.

An attempt was made to assess the clinical significance or social validation of changes (Kazdin, 1982) in these students by asking 'significant others' who are affected in important ways by the student's behaviour to assess changes (Kazdin, 1982). However, the problem with using 'significant others' is that this often does not have established reliability and validity (K. D. O'Leary & Turkewitz, 1978). Thus in addition these students' performance were compared with a few of their peers who were classified as 'normal'. With four of the students their teachers reported a noticeable improvement in their maths computational skills. Teacher reports for the other three students were unavailable. The interstudent comparisons indicate that the experimental students obtained accuracy levels comparable or higher than those of their 'normal' peers. Obviously, in future research, a systematic assessment of performance changes in the classroom would be desirable as well as evaluation of such changes over a follow-up period. Because this research was completed during the final weeks of the school year, such a follow-up assessment in the classroom was not feasible. A possible practical improvement concerns the issue of efficiency. In this research SI was taught on an individual basis, a process which would be time consuming for the normal classroom teacher. Although teachers normally do give individualised instruction to students, to implement the procedure suggested here a teacher would probably require the assistance of a teacher-aide. More cost effective would be to teach the children in a group, a procedure which has been shown to be successful (Whitman & Johnston, 1983).

Teachers need to pay more careful attention to errors than simply circling them or even just marking the problem wrong (and then just having students repeat the problems). G. H. Roberts (1968) concluded that the teacher who analyses the students' procedure when incorrectly solving a problem, will be in a better position to choose appropriate measures to help him/her overcome his/her particular difficulties and hence raise the level of competence. One important implication from the present study is that teachers should reinforce students for correct performance in a link. Sometimes an incorrectly solved problem is due to a small error in a particular link, for example, a calculation error. This was often the case for most students towards the end of this study. Often students in the classroom are not given reinforcement for those links which are correct, when the answer is incorrect due to a small error in a particular link. This is particularly necessary for students who have experienced failure often and who have become learned-helpless and unmotivated.

Error analysis could be carried out in future research to examine the possible relationships between error types and the checking procedures. For instance, is it is true that mc is more effective for errors in basic facts than algorithm errors? Furthermore, it would be interesting to measure the number of self-corrections made under each procedure. It is envisaged that this measure might help in interpreting the differential effects between the procedures during skill acquisition.

The procedures used in this study could also be usefully explored with younger children who would have had no prior training on long multiplication/division. Students in this study were teenagers/adolescents who may differ from younger children in several ways. First, students in this study were doing something which was not of interest (e.g., verbalisation) or compatible to their age. Research has shown that younger children with less pretraining experience appear to profit more from SI training when the verbalisations are more structured, detailed, and specific (Miller, Weinstein, & Karniol, 1978). Second, students here have longer history in experiencing failure and thus find maths learning aversive. In fact research has indicated that young children are considerably less vulnerable to the debilitating effects of failure than older children (Licht & Kistner, 1986). Third, students of different ages may differ in the amount of prior training on these types of problems, that is, teenagers have had some form of training on these problems.

Final mention should be made of the potential limitations due to the use of an AT design. As noted by Barlow and Hayes (1979), probably the greatest threat to the internal validity of this design is that of multiple treatment interference. Multiple treatment interference is "likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable" (Campbell & Stanley, 1963, p. 6). The present study took three steps to decrease the chances of multiple treatment interference. First, long intercomponent intervals between the treatment sessions were used (McGonigle, Rojahn, Dixon, & Strain, 1987). Second, systematic association of discriminative stimuli with conditions was used to make the discrimination between the conditions clear to the students. Third, treatments were randomised and counterbalanced. However, no direct assessment of such interference was made in the present study, although the reversal aspect of the end phases may have allowed for some examination of multiple treatment interference. It is conceivable, for example that "contrast" effects related to performing mc and ec were present for some students like Sue, but these effects were not strong. If such effects are present they would have a direct bearing on the validity of the obtained findings. Future research might address this issue by directly assessing the extent to which such effects are present using procedures suggested by Barlow and Hayes (1979).

There is disagreement as to what constitutes clear demonstration of experimental control when using an AT design. Clearest evidence of experimental control exists if the data paths under different treatments show no overlap. If there is overlap, the degree of experimental control demonstrated becomes arguable. Some would then consider that the best treatment can be identified only by showing that some statistical analysis of the differences between two paths reaches significance (Barlow & Hersen, 1984). Others would argue that experimental control can still be identified, although more weakly, if the majority of data points fall outside the range of values of the majority of data points for the contrasting treatments (e.g., Cooper et al., 1987). This latter usage was adopted in the present study as to identify the best treatment when experimental control was not clearly visible. The difficulty with using this criterion is that it has no specific rule on what constitutes a majority. However, in line with the major thrust of single subject research, one needs to make case by case arguments (J. M. Johnston & Pennypacker, 1980). Clearly in these instances further research must be done to support the findings.

One tactic that could have strengthened this study would be to also have a multiple baseline across subjects with the alternating treatments design. This would control for potential confounds that are a function of multiple measures and other conditions presented during baseline (Barlow & Hersen, 1984). The use of this feature was not added, again due to the practical limitation of time imposed by the school.

CONCLUSION

The present study was one of the few mathematics studies in the error-correction literature. It is the first to investigate error-correction techniques which are student-initiated. It is also the first to investigate the relative efficacy of two types of checking procedures and a no-checking procedure on maths accuracy. To date no study has examined this form of the self-evaluation component of an SI package. Unlike in most previous studies, an attempt was made in the present study to interpret the results in consistent behavioural terms.

There were several strengths to this study. The addition of a baseline phase prior to the introduction of the AT phase allowed further identification of the naturally occurring accuracy of the target problems and the increase in accuracy in the target problems when the treatments were instigated. Although this is not necessary to determine which of the three treatments was most effective, it provided additional information. Second, all students did better under mc, except Keith who did better under ec. This is a good example of handling intersubject variability in a single-subject design. A between-group design would average out, rather than highlight these individual differences in response to treatment. Because of this intersubject variability, the investigator needed to speculate on the likely explanations. Third, the reversal phases at the end of the study allowed the strengthening of the conclusion that mc was more effective than nc.

In spite of a number of shortcomings several clear conclusions can be drawn from the present study. First specific (verbalised) self-instructions in the form of a strategy substantially improved accuracy in multiplication and division problems irrespective of the procedures used. Second, mc was more effective than nc for all students (except Keith for whom ec was more effective than nc). Third, mc was typically more effective than ec for most students. Fourth, when a task of medium complexity was used for instruction, generalisation to problems of both greater and less complexity occurred.

A possible advantage of SI training over traditional didactic procedures merits attention. With traditional teaching methods, maths proficiency is typically assessed by performance scores yielding dichotomous information, that is, either the child does or does not solve the maths problem in question. When a student fails to master a skill, the teacher does not have immediately at his/her disposal accurate and useful information regarding the student's difficulty. In contrast, teaching through the SI approach provides the teacher with ongoing feedback about difficulties the student may be experiencing. Having the child verbalise the self-instructions as he/she works on a problem allows the teacher to quickly diagnose the student's difficulty and to focus remediation efforts on the specific point of difficulty.

Moreover, SI training has the advantage over other training modalities of being able to be tailored to the needs of individual students and to the attainment of specific tasks. It maintains control by the student and focuses on the process as well as the product of the problem solving procedure. Recently a number of behaviour therapists have emphasised the need to switch control of an individual's behaviour from external agents (e.g., the teacher) to the individual himself (e.g., Kazdin, 1975).

Emphasising the behavioural nature of private speech which, like any other behaviour, is affected by antecedents and consequences should result in a better understanding of the functional relationships involved, and, hence lead to better techniques of control. For instance, in the course of describing their procedure for establishing SI behaviour, CBM theorists advise that the meaning of one's self-statements should be internalised. But how does one do this? A behavioural approach to this problem would be to ensure that the student's statements functioned as effective discriminative stimuli for specific behaviours. Thus rather than simply reinforcing the student's self-statements, only the appropriate verbal behaviour (S^D)-other behaviour relationship would be reinforced. Evidence that such a procedure is effective comes from recent behavioural research on verbal-nonverbal correspondence (e.g., Guevremont, Osnes, & Stokes, 1986; Williams & Stokes, 1982).

The results of this study have several ramifications for future research. Clearly systematic replication is needed to test some of the hypotheses suggested in the present study which included the following: First, the effectiveness of mc depends on task or chain complexity, student's distractibility, preskill knowledge and also speed and stage of skill acquisition. Second, the differential effect of the procedures seems to be minimal once students become more proficient. Third mc may be effective only for detecting certain type of errors.

Further research on SI training is needed to (a) demonstrate the applied significance of checking procedures; (b) do a component analysis of the effects of each step in the SI package; (c) identify those behaviours, setting and populations for which SI training and error-monitoring are most appropriate; and (d) identify those variables that promote generalisation (of SI error-monitoring procedures).

This study has several important implications for teaching. First, it is possible for children with learning disabilities to learn academic skills if they are given adequate and specific instructions. Second, while they are learning a skill it is better for students to check as they go (multi-check) than to check after completion of the problem (end-check). Third, end-checking, the most commonly used procedure in schools may be no more effective than not checking once the student becomes proficient. Fourth, generalisation may readily occur depending on the type of academic instruction provided.

References

Albion, F. M., & Salzberg, C. L. (1982). The effect of self-instructions on the rate of correct addition problems with mentally retarded children. *Education and Treatment of Children, 5,* 121-131.

American Psychological Association (1983). *Publication Manual of the American Psychological Association* (3rd ed.). Washington, DC: Author

Baer, D. M. (1982). Applied behavior analysis. In G. T. Wilson & C. M. Franks (Eds.), *Contemporary behavior therapy* (pp. 277-309). New York: The Guildford Press.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91-97.

Baer, R. A., Blount, R. L., Detrich, R., & Stokes, T. F. (1987). Using intermittent reinforcement to program maintenance of verbal nonverbal correspondence. *Journal of Applied Behavior Analysis, 20,* 179-184.

Baine, D. (1982). *Instructional design for special education.* Englewood Cliffs, NJ: Educational Technology Publications.

Barkley, R. A., Copeland, A. P., & Sivage, C. (1980). A self-control classroom for hyperactive children. *Journal of Autism and Developmental Disorders, 10,* 75-89.

Barling, J. (1980). A multistage multidependent variable assessment of children's self regulation of academic performance. *Cognitive Behavior Therapy, 2,* 43-54.

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12,* 199-210.

Barlow, D. H., Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.

Becker, W. C., & Carnine, D. W. (1981). Direct instruction: A behavior theory model for comprehensive educational intervention with the disadvantaged. In S. W. Bijou & R. Ruis (Eds.), *Behavior Modification: Contributions to education.* Hillsdale, NJ: Erlbaum.

Brueckner, L. J., & Melbye, H. O. (1940). Relative difficulty of types of examples in division with two-figure divisors. *Journal of Educational Research, 33,* 401-414.

Blankenship, C. S. (1978). Remediating systematic inversion errors in subtraction through use of demonstration and feedback. *Learning Disability Quarterly, 1,* 12-22.

Blankenship, C. S., & Baumgartner, M. D. (1982). Programming generalization of computational skills. *Learning Disability Quarterly, 5,* 152-162.

Blankenship, C. S., & Lilly, M. S. (1981). *Mainstreaming students with learning problems: Techniques for the classroom teacher.* New York: Holt, Rinehart & Winston.

Brigham, T. A. (1980). Self-control revisited: Or why doesn't anyone actually read Skinner (1953). *The Behavior Analyst, 3,* 25-33.

Broughton, S. F., & Lahey, B. B. (1978). Direct and collateral effects of positive reinforcement, response cost and mixed contingencies for academic performance. *Journal of School Psychology, 16*, 126-136.

Burgio, L. D., Whitman, T. L., & Johnson, M. R. (1980). A self-instructional package for increasing attending behavior in educable mentally retarded children. *Journal of Applied Behavior Analysis, 13*, 443-459.

Bursuck, W. D., Epstein, M. H. (1987). Current research topics in learning disabilities. *Learning Disability Quarterly, 10(1)*, 2-7.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In D. T. Campbell and J. C. Stanley, *Handbook of Research on Teaching*. Chicago: Rand McNally.

Carnine, D. (1983). Direct Instruction: In search of instructional solutions for educational problems. In *Interdisciplinary voices in learning disabilities and remedial education* (pp. 1-66). Austin, TX: Pro-Ed.

Cawley, J. F., Fitzmaurice, A. M., Shaw, R. A., Kahn, H., & Bates, H. (1978). Mathematics and learning disabled youth: The upper grades. *Learning Disability Quarterly, 1(4)*, 37-52.

Chapman, J. W. (1985). *Self-perceptions of ability, learned helplessness and academic achievement expectations of children with learning disabilities* (Contract No. 55-2-18). Palmerston North, NZ: Massey University, Department of Education Research.

Chapman, J. W., & van Kraayenood, C. E. (1987). Learning disabilities. In D. Mitchell & N. Singh (Eds.), *Exceptional Children in New Zealand*. Palmerston North, NZ: Dunmore Press.

Chapman, J. W.,& Wilkinson, L. (1988). *Self-Concept, identity status and emotional status of form 5 learning disabled students: A longitudinal study of intermediate school predictors* (Contract No. 47/17/151). Palmerston North: Massey University, Department of Education Research.

Chiang, B. (1986). Initial learning and transfer effects of microcomputer drills on LD students' multiplication skills. *Learning Disability Quarterly, 9(2)*, 118-123.

Cohen, R., Rubin, S., & Heinen, J. (1979). Collateral increases in performance of long division resulting from on-task gains using token reinforcement and self recording. *Psychological Reports, 44*, 651-658.

Cole, P. M., & Kazdin, A. E. (1980). Critical issues in self-instructional training with children. *Child Behavior Therapy, 2*, 1-23.

Conlon, M. F., Hall, C., & Hanley, E. M. (1972). The effects of a peer correction procedure on the arithmetic accuracy for two elementary school children. In G. Semb (Ed.), *Behavior Analysis and Education* (pp. 205-210). Lawrence, KS: Univ of Kansas Support and Development for Follow-Through.

Conners, C. K. (1969). A Teacher Rating Scale for use in drug studies with children. *American Journal of Psychiatry, 126*, 152-156.

Conners, C. K. (1973). Rating scales for use in drug studies with children. *Psychopharmacology Bulletin* (special issue, Pharmacotherapy of Children), 24-29.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis.* Columbus, OH: Merrill Publishing Company.

Cox, L. S. (1975a). Diagnosing and remediating systematic errors in addition and subtraction computations. *The Arithmetic Teacher, 22*, 151-157.

Cox, L. S. (1975b). Systematic errors in the 4 algorithms in normal and handicapped populations. *Journal for Research in Mathematics Education, 6*, 202-220.

Cullinan, D., Epstein, M. H., Lloyd, J. (1978). *The effects of self-verbalization on the accuracy of multiplication performance.* Paper presented at the annual convention of the association for children with learning disabilities, Kansas City.

Cullinan, D., Lloyd, J., & Epstein, M. H. (1981). Strategy training: A structured approach to arithmetic instruction. *Exceptional Education Quarterly, 2(1)*, 41-49.

Deguchi, H. (1984). Observational learning from a radical-behavioristic viewpoint. *The Behavior Analyst, 7*, 83-95.

Dolly, J. P. (1980). The role of stimulus control in instructional design for learning disabled children. *Journal for Special Educators, 16*, 235-242.

Drucker, H., McBridge, S., & Wilbur, C. (1987). Using a computer-based error analysis approach to improve basic subtraction skills in the third grade. *Journal of Educational Research, 80*, 363-365.

Elley, W. B., & Reid, N. A. (1969). *Progressive Achievement Tests: Reading Comprehension & Reading Vocabulary.* Wellington: New Zealand Council for Educational Research.

Elley, W. B., & Reid, N. A. (1971). *Progressive Achievement Tests: Listening Comprehension.* Wellington: New Zealand Council for Educational Research.

Farrald, R. R., & Schamber, R. G. (1973). *A diagnostic and prescriptive techniqe: Handbook 1: A mainstream approach to the identification, assessment, and amelioration of learning disabilities.* Sioux Falls, SD: Adapt Press.

Fink, W. T., & Carnine, D. W. (1975). Control of arithmetic errors using informational feedback and graphing. *Journal of Applied Behavior Analysis, 8*, 461.

Fleischner, J. E., & Garnett, K., & Shepard, M. J. (1982). Proficiency in arithmetic basic facts computation of learning disabled and nondisabled children. *Focus on Learning Problems in Mathematics, 4(2)*, 47-56.

Forbes, S. D. (1987). Mathematical priorities in secondary education. *The New Zealand Mathematical society newsletter, 40*, 23-30.

Forness, S. R., & MacMillan, D. L. (1970). The origins of behavior modification with exceptional children. *Exceptional Children, 37*, 93-100.

Friedling, C., & O'Leary, S. G. (1979). Effects of self- instructional training on second- and third-grade hyperactive children: A failure to replicate. *Journal of Applied Behavior Analysis, 12*, 211-219.

Gadow, K. D., Torgesen, J. K., & Dahlem, W. E. (1983). Learning Disabilities. In M. Hersen, V. B. Hasselt, & J. L. Matson (Eds.), *Behavior Therapy for the developmentally and physically disabled: A handbook* (pp 310-351). New York: Academic Press.

Gerber, M. M. (1987). Application of cognitive-behavioral training methods to teaching basic skills to mildly handicapped elementary school students. In M. C. Wang, M. C. Reynolds, and H. J. Walberg (Eds.). *Handbook of special education research and practice: Volume 1. Learner characteristics and adaptive education* (pp. 167-186). Oxford: Pergamon Press.

Graden, G., Thurlow, M., & Ysseldyke, J. (1983). Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence. *Journal of Experimental Child Psychology, 36*, 241-256.

Gresham, F. M. (1988). Social competence and motivational characteristics of learning disabled students. In M. C. Wang, M. C. Reynolds, and H. J. Walberg (Eds.). *Handbook of special education research and practice: Volume 2. Mildly handicapped conditions* (pp. 283-302). Oxford: Pergamon Press.

Grimes, L. (1981). Error analysis and error correction procedures. *Teaching Exceptional Children, 14*, 17-20.

Grimm, J., Bijou, S., & Parsons, J. (1973). A problem solving model for teaching remedial arithmetic to handicapped young children. *Journal of Abnormal Child Psychology, 7*, 26-39.

Grossnickle, F. E. (1936). Errors and questionable habits of work in long division with a one-figure divisor. *Journal of Educational Research, 29*, 355-368.

Grossnickle, F. E. (1939). Constancy of error in learning division with a two-figure divisor. *Journal of Educational Research, 33*, 189-196.

Guevremont, D. C., Osnes, P. G., & Stokes, T. F. (1986). Programming maintenance after correspondence training interventions with children. *Journal of Applied Behavior Analysis, 19*, 215-219.

Guilford, J. P. (1965). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.

Hallahan, D. P., & Kauffman, J. M. (1976). *Introduction to learning disabilities: A psychobehavioral approach*. Englewood Cliffs, NJ.: Prentice Hall.

Hallahan, D. P., & Kauffman, J. M. (1977). Labels, categories, behaviors: ED, LD, and EMR reconsidered. *Journal of Special Education, 11*, 139-149.

Hallahan, D. P., Kneedler, R. D., & Lloyd, J. W. (1983). Cognitive behavior modification techniques for learning disabled children: Self-instruction and self-monitoring. In J. D. McKinney & L. Feagans (Eds.), *Current Topics in Learning Disabilities*, Vol 1. New York: Ablex Publishing.

Hallahan, D. P., Lloyd, J. W., Kneedler, R. D., & Marshall, K. J. (1982). A comparison of the effects of self versus teacher assessment of on-task behavior. *Behavior Therapy, 13*, 715-723.

Hallahan, D. P., Lloyd, J. W., Kosiewicz, M. M., Kauffman, J. M., & Graves, A. W. (1979). Self monitoring of attention as a treatment for a learning disabled boy's off-task behavior. *Learning Disability Quarterly, 2(3)*, 24-32.

Hartmann, D. P. (1977). Consideration in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis, 10*, 527-532.

Hartig, M., & Kanfer, F. H. (1973). The role of verbal self-instructions in children's resistance to temptation. *Journal of Personality and Social Psychology, 25*, 259-267.

Hasazi, J. E., & Hasazi, S. E. (1972). Effects of teacher attention on digit reversal behaviors in an elementary school child. *Journal of Applied Behavior Analysis, 5*, 157-162.

Heins, E. D. (1980). Training learning disabled children's self-control: Cued and non-cued self-recording in the classroom (Doctoral dissertation, University of Virginia). *Dissertation Abstracts International, 41*, 3986A.

Hobbs, S. A., & Lahey, B. B. (1977). The behavioral approach to "Learning Disabled" Children. *Journal of Clinical Child Psychology, 6*, 10-14.

Howell, R., Sidorenko, E., & Jurica, J. (1987). The effects of computer use on the acquisition of multiplication facts by a student with learning disabilities. *Journal of Learning Disabilities, 20*, 336-341.

Hundert, J., & Bucher, B. (1978). Pupils self-scored arithmetic performance: A practical, procedure for maintaining accuracy. *Journal of Applied Behavior Analysis, 11*, 304.

Israel, A. C. (1973). Developing correspondence between verbal and nonverbal behavior: Switching sequences. *Psychological Reports, 32*, 1111-1117.

Israel, A. C. (1978). Some thoughts on correspondence between saying and doing. *Journal of Applied Behavior Analysis, 11*, 271-276.

Israel, A. C., & O'Leary, K. D. (1973). Developing correspondence between children's words and deeds. *Child Development, 44*, 575-581.

Jenkins, J. R., & Gorrafa, S. (1972). Superimposing contracts upon a token economy. In G. Semb (Ed.), *Behavior analysis and education* (pp. 36-368). Lawrence, KS: Univ of Kansas Support and Development for Follow-Through.

Jenkins. J. R., & Larson, L. (1979). Evaluating error-correction procedures for oral reading. *Journal of Special Education, 13*, 145-156.

Johnston, R. J., & McLaughlin, T. F. (1982). The effects of free time on assignment completion and accuracy in arithmetic: A case study. *Education and Treatment of Children, 5*, 33-40.

Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnston, M. B., Whitman, T. L., & Johnson, M. (1980). Teaching addition and subtraction to mentally retarded children : A self-instruction program. *Applied Research in Mental Retardation, 1*, 141-160.

Karoly, P., & Dirks, M. J. (1977). Developing self-control in pre-school children through correspondence training. *Behavior Therapy, 8*, 398-405.

Kaufman, A. S. (1976). A new approach to the interpretation of test scatter on the WISC-R. *Journal of Learning Disabilities, 9*, 160-168.

Kauffman, J. M. (1975). Behavior modification. In W. M. Cruickshank & D. P. Hallahan (Eds.), *Perceptual and learning disabilities in children.* Syracuse, N.Y.: Syracuse Univ. Press.

Kauffman, J. M. (1981). *Characteristics of children's behavior disorders* (2nd ed.). Columbus, OH: Charles E. Merrill.

Kavale, K. A., Forness, S. R. (1985). *The science of learning disabilities.* San Diego, CA: College Hill Press.

Kavale, K., & Nye, C. (1981). Identification criteria for learning disabilities: A survey of the research literature. *Learning Disability Quarterly, 4,* 383-388.

Kazdin, A. E. (1975). *Behavior modification in applied settings.* Homewood, Ill: The Dorsey Press.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford Univ. Press.

Kirby, F. D., & Shields, F. (1972). Modification of arithmetic response rate and attending behavior in a 7th grade student. *Journal of Applied Behavior Analysis, 5,* 78-84.

Koorland, M. A. (1986). Applied behavior analysis and the correction of learning disabilities. In J. K. Torgesen & B. Y. L. Wong (Eds.), *Psychological and educational perspectives on learning disabilities (pp. 297-328).* Orlando: Academic Press.

Leon, J. A., & Pepe, H. J. (1983). Self -instructional training: Cognitive behavior modification for remediating arithmetic deficits. *Exceptional Children, 50(1),* 54-61.

Licht, B. G., & Kistner, J. A. (1986). Motivational problems of learning-disabled children: Individual differences and their implications for treatment. In J. K. Torgesen & B. Y. L. Wong (Eds.), *Psychological and educational perspectives on learning disabilities (pp. 225-255).* Orlando: Academic Press.

Lloyd, J. W. (1980). Academic instruction and cognitive-behavior modification: The need for attack strategy training. *Exceptional Education Quarterly, 1(1),* 53-63.

Lloyd, J. W. (1988). Direct academic interventions in learning disabilities. In M. C. Wang, M. C. Reynolds, and H. J. Walberg (Eds.), *Handbook of special education research and practice: Volume 2. Mildly handicapped conditions* (pp. 345-366). Oxford: Pergamon Press.

Lloyd, J., Hallahan, D. P., Kosiewicz, M. M., & Kneedler, R. D. (1982). Reactive effects of self-assessment and self-recording on attention to task and academic productivity. *Learning Disability Quarterly, 5,* 216-227.

Lloyd, J., Saltzman, N. J., & Kauffman, J. M. (1981). Predictable generalization in academic learning as a result of preskills and strategy learning. *Learning Disability Quarterly, 4,* 203-256.

Lovitt, T. C. (1975a). Part 1: Characteristics of ABA, general recommendations, and methodological limitations. *Journal of Learning Disabilities, 8,* 432-443.

Lovitt, T. C. (1975b). Part 2: Specific research recommendations and suggestions for practitioners. *Journal of Learning Disabilities, 8,* 504-518.

Lovitt, T. C., & Curtiss, K. A. (1968). Effects of manipulating antecedent events on math response rate. *Journal of Applied Behavior Analysis, 1,* 329-333.

Lovitt, T. C., & Esveldt, K. A. (1970). The relative effects on math performance of single- versus multiple-ratio schedules: a case study. *Journal of Applied Behavior Analysis, 3,* 261-272.

Lowe, C. F. (1979). Determinants of human operant behavior. In M. D. Zeiler and P. Harzem (Eds.), *Advances in analysis of behavior: Volume 1. Reinforcement and the organisation of behavior.* Chichester, England: Wiley.

Lowe, C. F., Harzem, P., & Bagshaw, M. (1978). Species differences in temporal control of behavior II: human performance. *Journal of Experimental Analysis of Behavior 29,* 351-61.

Lowe, C. F., Harzem, P., & Hughes, S. (1978). Determinants of operant behavior in humans: some differences from animals. *Journal of Experimental Psychology, 30,* 373-386.

Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior.* New York: Liverwight.

McLeod, T. M., & Armstrong, S. W. (1982). Learning disabilities in mathematics: Skill deficits and remedial approaches at the intermediate and secondary level. *Learning Disability Quarterly, 5,* 305-311.

McNeil, M., Hasazi, S., Muller, A., & Knight, M. (1972). Open classrooms: Supporters of applied behavior analysis. In G. Semb (Ed.), *Behavior analysis and education* (pp. 112-126). Lawrence, KS: Univ of Kansas Support and Development for Follow-Through .

McGonigle, J. J., Rojahn, J., Dixon, J., & Strain, P. S. (1987). Multiple treatment interference in the alternating treatments design as a function of the intercomponent interval length. *Journal of Applied Behavior Analysis, 20,* 171-178.

Meichenbaum, D., & Goodman, J. (1969). Reflection-impulsivity and verbal control of motor behavior. *Child Development, 40,* 785-797.

Meichenbaum, D., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self control. *Journal of Abnormal Psychology, 77,* 115-126.

Mercer, C. D. (1979). *Children and adolescents with learning disabilities.* Columbus, OH: Charles E. Merrill.

Mercer, C. D. (1983). *Students with learning disabilities* (2nd ed.). Columbus, OH: Charles E. Merril.

Miller, D. T., Weinstein, S. M., & Karniol, R. (1978). Effects of age and self-verbalization on children's ability to delay gratification. *Developmental Psychology, 11,* 569-570.

Mitchell, I. V. (Ed.) (1985). *The Ninth Mental Measurements Year book.* The Buros Institute of Mental Measurement: University of Nebraska Press.

Montague, M., & Bos, C. S. (1986). The effect of cognitive strategy training on verbal math problem solving performance of learning disabled adolescents. *Journal of Learning Disabilties, 19(1),* 26-33.

Moon, B. (1986). *The new maths' curriculum controversy.* London: Falmer press.

Moore, J. (1980). On behaviorism and private events. *The Psychological Record, 30,* 459-475.

Moore, J. (1984). On behaviorism, knowledge, and causal explanation. *The Psychological Record, 34,* 73-97.

Morsink, C. V., Thomas, C. C., & Smith-Davis, J. (1987). Noncategorical special education programs: Process and outcomes. In M. C. Wang, M. C. Reynolds, and H. J. Walberg (Eds.). *Handbook of special education research and practice: Volume 1. Learner characteristics and adaptive education* (pp. 287-309). Oxford: Pergamon Press.

Nelson, R. O., & Hayes, S. C. (1981). Theoretical explanations for reactivity in self-monitoring. *Behavior Modification, 5*, 3-14.

O'Leary, S. G., & Dubey, D. R. (1979). Applications of self-control procedures by children: A review. *Journal of Applied Behavior Analysis, 12*, 449-465.

O'Leary, K. D., & O'Leary, S. G. (1977). *Classroom management*. New York: Pergamon.

O'Leary, K. D., & Turkewitz, H. (1978). Methodological errors in marital and child treatment research. *Journal of Consulting and Clinical Psychology, 46*, 747-758.

Paine, S. C., Carnine, D. W., White, W. A. T., & Walters, G. (1982). Effects of fading teacher presentation structure (covertization) on acquisition and maintenance of arithmetic problem-solving skills. *Education and Treatment of children, 5*, 93-107.

Palkes, H., Stewart, M., & Freedman, J. (1972). Improvement in maze performance of hyperactive boys as a function of verbal-training procedures. *The Journal of Special Education, 5*, 337-342.

Parsons, J. A. (1972). The reciprocal modification of arithmetic behavior and program development. In G. Semb (Ed.), *Behavior analysis and education (pp. 185-199).* Lawrence, K.S: Univ. of Kansas Support and Development for Follow-Through.

Pierson, P. M. (1972). Three student studies. In G. Semb (Ed.), *Behavior analysis and education* (pp. 290-300). Lawrence, KS: Univ of Kansas Support and Development for Follow-Through.

Pollack, M. J., Sulzer-Azaroff, B., & Williams, R. (1972). The experimental analysis of a homeward educational program with a pre-delinquent juvenile. In G. Semb (Ed.), *Behavior analysis and education* (pp. 303-309). Lawrence, KS: Univ of Kansas Support and Development for Follow-Through.

Reid, N. A., & Hughes, D. C. (1974). *Progressive Achievement Tests: Mathematics.* Wellington: New Zealand Council for Educational Research.

Reisman, F. K. (1982). *A guide to the diagnostic teaching of arithmetic.* Columbus, OH: Charles E. Merrill.

Reschly, D. J. (1987). Learning characteristics of mildly handicapped students: Implications for classification, placement, and programming. In M. C. Wang, M. C. Reynolds, and H. J. Walberg (Eds.). *Handbook of special education research and practice: Volume 1. Learner characteristics and adaptive education* (pp. 35-58). Oxford: Pergamon Press.

Reynolds, C. R. (1982). The problem of bias in psychological assessment. In C. R. Reynolds and T. B. Gutkin (Eds.), *Handbook of School Psychology.* New York: John Wiley.

Risley, T. R. (1977). The social context of self-control. In R. B. Stuart (Ed.), *Behavioral self-management: Strategies, techniques and outcomes.* New York: Brunner/Mazel.

Rivera, D. M., & Smith, D. D. (1987). The influence of modeling on acquisition and generalization of computational skills: A summary of findings from three sites. *Learning Disability Quarterly, 10(1)*, 69-80.

Rivera, D., & Smith, D. D. (1988). Using a demonstration strategy to teach midschool students with learning disabilities how to compute long division. *Journal of Learning Disabilities, 21*, 77-81.

Roberts, G. H. (1968). The failure strategies of third grade arithmetic pupils. *The Arithmetic Teacher, 15*, 442-446.

Roberts, R. N. (1979). Private speech in academic problem solving: A naturalistic perspective. In G. Zivin (Ed.), *The development of self-regulation through private speech*. New York: Wiley.

Roberts, R. N., & Nelson, R. O. (1983). Assessment issues and strategies in cognitive behavior therapy with children. *Cognitive Behavior Therapy, 5*, 99-128.

Roberts, R. N., Nelson, R. O., & Olson, T. W. (1987). Self-instruction: An analysis of the differential effects of instruction and reinforcement. *Journal of Applied Behavior Analysis, 20*, 235-242.

Rogers-Warren, A. R., & Baer, D. M. (1976). Correspondence between saying and doing: teaching children to share and praise. *Journal of Applied Behavior Analysis, 9*, 335-354.

Rooney, K., Polloway, E. A., & Hallahan, D. P. (1985). The use of self-monitoring procedures with low IQ learning disabled students. *Journal of Learning Disabilities, 18*, 384-389.

Rose, R., Koorland, M., & Epstein, J. (1982). A review of applied behavior analysis interventions with learning disabled children. *Education and Treatment of Children, 5*, 41-58.

Rose, T. L., McEntire, E., & Dowdy, C. (1982). Effects of two error-correction procedures on oral reading. *Learning Disability Quarterly, 5*, 100-105.

Rosenberg, M. S. (1986). Error correction during oral reading – A comparison of 3 techniques. *Learning Disability Quarterly, 9(3)*, 182-192.

Rosenberg, M. S. (1989). The effects of daily homework assignments on the acquisition of basic skills by students with learning disabilities. *Journal of Learning Disabilities, 22*, 314-323.

Ryan, E. B., Weed, K. A., & Short, E. J. (1986). Cognitive-behavior modification: Promoting active, self-regulatory learning styles. In J. K. Torgesen & B. Y. L. Wong (Eds.), *Psychological and educational perspectives on learning disabilities (pp. 367-397)*. Orlando, FL: Academic Press.

Salzberg, C. L. (1972). Freedom and responsibility in an elementary school. In G. Semb (Ed.), *Behavior analysis and education* (pp. 62-77). Lawrence, K.S: Univ of Kansas Support and Development for Follow-Through .

Schloss, P. J., Sedlak, R. A., Elliott, C., & Smothers, M. (1982). Application of the changing-criterion design in special education. *The Journal of Special Education, 16*, 359-367.

Schumaker, J., & Sherman, J. A. (1970). Training generative verb usage by imitation and reinforcement procedures. *Journal of Applied Behavior Analysis, 3*, 273-287.

Schweitzer, J. B., & Sulzer-Azaroff, B. (1988). Self-control: Teaching tolerance for delay in impulsive children. *Journal of Experimental Analysis of Behavior, 50(2)*, 173-186.

Shepard, L. A., Smith, M. L., & Vojir, C. P. (1983). Characteristics of pupils identified as learning disabled. *American Educational Research Journal, 20*, 309-331.

Singh, N. N., Winton, A. S., & Singh, J. (1985). Effects of delayed versus immediate attention to oral reading errors on the reading proficiency of mentally retarded children. *Applied Research in Mental Retardation, 6*, 283-293.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review, 52*, 270-277.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Skinner, B. F. (1957). *Verbal behavior.* London: Methuen.

Skinner, B. F. (1963). Behaviorism at fifty. *Science, 134*, 566-602.

Skinner, B. F. (1974). *About behaviorism.* New York: Alfred A. Knopf.

Smith, D. D., & Lovitt, T. C. (1975). The use of modeling techniques to influence the acquisition of computational arithmetic skills in learning-disabled children. In E. Ramp & G. Semb (Eds.), *Behavior analysis: Areas of research and application* (pp. 283-308). Englewood Cliffs, NJ: Prentice-Hall.

Smith, D. D., & Lovitt, T. C. (1976). The differential effects of reinforcement contingencies on arithmetic performance. *Journal of Learning Disabilities, 9*, 21-29.

Smith, D. D., Lovitt, T. C., & Kidder, J. D. (1972). Using reinforcement contingencies and teaching-aids to alter subtraction performance of children with learning disabilities. In G. Semb (Ed.), *Behavior analysis and education* (pp. 342-360). Lawrence, KS: Univ. of Kansas Support and Development for Follow-Through.

Snider, V. (1987). Use of self-monitoring of attention with LD students: Research and application. *Learning Disability Quarterly, 10*, 139-151.

Speltz, M. L., Shimamura, J. W. & McReynolds, W. T. (1982). Procedural variations in group contingencies: Effects on academic and social behaviors. *Journal of Applied Behavior Analysis, 15*, 533-544.

Sprague, R. L., Cohen, M., & Werry, J. S. (1974). *Normative data on the Conners Teacher Rating Scale and Abbreviated Scale* (Technical Report). Illinois: University of Illinois, Children's Research Center.

Stokes, T., & Baer, D. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349-367.

Sulzer-Azaroff, B., & Mayer, G. R. (1977). *Applying behavior analysis procedures with children and youth.* New York: Holt.

Taylor, R. L. (1984). *Assessment of exceptional students: Educational and Psychological Procedures.* Englewood Cliffs, NJ: Prentice-Hall.

Throne, J. M. (1973). Learning disabilities: A radical behaviorist point of view. *Journal of Learning Disabilities, 6*, 543-546.

Treiber, F. A., & Lahey, B. B. (1983). Toward a behavioral model of academic remediation with learning disabled children. *Journal of Learning Disabilities, 16*, 111-116.

Van Houten, R., & Rolider, A. (1989). An analysis of several variables influencing the efficacy of flash card instruction. *Journal of Applied Behavior Analysis, 22*, 111-118.

Van Houten, R., & Thompson, C. (1976). The effects of explicit timing on mathematics performance. *Journal of Applied Behavior Analysis, 9*, 227-230.

Van Luit, J., & Van der Aalsvoort, G. (1985). Learning subtraction in a special school: a self instructional training strategy for educable mentally retarded children with arithmetic deficits. *Instructional Science, 14*, 179-189.

Varni, J. W., & Henker, B. (1979). A self-regulation approach to the treatment of three hyperactive boys. *Child Behavior Therapy, 1*, 171-192.

Vygotsky, L. (1934). *Thought and language.* New York: Wiley, 1962.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children - Revised.* New York: The Psychological Corporation.

White, O. R., & Haring, N. G. (1976). *Exceptional Teaching.* Columbus, OH: Charles E. Merril.

Whitman, T. L., & Johnston, M. B. (1983). Teaching addition and subtraction with regrouping to educable mentally retarded children: A group self-instructional training program. *Behavior Therapy, 14*, 127-143.

Wilder, L., Draper, T. W., & Donnelly, C. P. (1984). Overt and covert verbalization in normal and learning disabled children's problem solving. *Perceptual and Motor Skills, 58*, 976-978.

Williams, J. A., & Stokes, T. F. (1982). Some parameters of correspondence training and generalized verbal control. *Child and Family Behavior Therapy, 4*, 11-31.

Wong, B. Y. L. (1986). Problems and issues in the definition of learning disabilities. In J. K. Torgesen & B. Y. L. Wong (Eds.), *Psychological and educational perspectives on learning disabilities (pp. 3-26).* Orlando: Academic Press.

Ysseldyke, J. E., Algozzine, B., Shinn, M. & McGue, M. (1982). Similarities and differences between low achievers and students classified learning disabled. *The Journal of Special Education, 16*, 73-85.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. (1983). On interpreting institute research : A response to McKinney. *Exceptional Child Quarterly, 4*, 145-147.

Appendix 1

SAMPLE MULTIPLICATION AND DIVISION WORKSHEETS

WORKSHEET 11A

Name:________________________ Date:____________

Attempt as many problems as you can. You have 20 minutes.

1) 7083 X 3

2) 6438 X 25

3) 32567 X 7

4) 9449 X 5

5) 2026 X 454

6) 6636 X 456

7) 8050 X 5

8) 152 X 45

9) 6406 X 237

10) 43 X 74

11) 936 X 9

12) 547 X 511

13)
```
  4 0 6 7
X   9 8 6
```

14)
```
  4 3 9 5
X     4 2
```

15)
```
  5 1 7 5
X       4
```

16)
```
   3 4 5
X  3 6 2
```

17)
```
  5 0 0 6
X   7 4 5
```

18)
```
  6 8 5
X     8
```

19)
```
  6 0 0 7
X     4 3
```

20)
```
  7 5 0 6
X       2
```

21)
```
  7 4
X 1 6
```

22)
```
  3 4 4 5 2
X         5
```

23)
```
  7 0 6
X   3 5
```

24)
```
  5 1 7 6
X   3 6 8
```

WORKSHEET 13B

Name:________________________ Date:____________

Attempt as many problems as you can. You have 20 minutes.

1) $7\overline{)4904}$

2) $31\overline{)56918}$

3) $3\overline{)618}$

4) $325\overline{)6175}$

5) $33\overline{)6164}$

6) $16\overline{)64608}$

7) $223\overline{)69174}$

8) $381\overline{)88777}$

9) $4\overline{)496}$

10) $44\overline{)528}$

11) $374\overline{)34780}$

12) $7\overline{)14835}$

13) $2\overline{)7227}$

14) $367\overline{)89551}$

15) $21\overline{)699}$

16) $46\overline{)36940}$

17) $8\overline{)64018}$

18) $281\overline{)64916}$

19) $4\overline{)369}$

20) $16\overline{)36425}$

21) $234\overline{)5616}$

22) $38\overline{)1976}$

23) $111\overline{)98799}$

24) $6\overline{)24521}$

Appendix 2

SAMPLE SCRIPTS USED DURING TRAINING OF MULTIPLICATION AND DIVISION

The general rules used in solving a double-digit multiplier/divisor problem is presented in bold. The application of each general rule for the particular problem is indicated within < >, followed by the detailed steps which are enclosed in * *. Details have not been repeated in areas where it is obvious, for example, in obtaining the some of the partial dividends while computing long division.

SAMPLE SCRIPT FOR THE MULTIPLICATION TRAINING SESSION

Step 1
Experimenter

Demonstrates the first checking procedure (e.g., end-checking)
problem: 6807 x 65

"This is the procedure where we check at the end."

"We first multiply the multiplicand by the units digit of the multiplier.
< We first multiply 6807 by 5 ones >
* 5x7 is 35, 5, carry 3 above the 0; 5x0 is 0, plus 3 is 3; 5x8=40, 0 carry 4; 5x6=30, plus 4, equals 34 *

Now we multiply the multiplicand by the tens digit of the multiplier. So put a zero and shift numerals one place to the left, that is, indent one place to the left.
< Now we multiply 6807 by 6 tens so we put a zero and move one place to the left >
* Place one zero. 6x7=42, 2 carry 4; 6x0=0, plus 4 is 4; 6x8=48, 8 carry 4; 6x6=36, plus 4 is 40; write 0 then 4 *

Add the partial products
< I add the columns >
* 5+0=5; 3+2=5; 0+4=4; 4+8=12, 2 carry 1; 1+3+0=4; and 0+4=4 *

Now to end-check:
Check structure (Have a zero for the second partial product and moved one place to the left?),
< I'm multiplying by a tens digit, so do I have a zero in the 2nd partial product? – yes >

Check alignment, and then check multiplication (need not check all the multiplication links),
< Aligned numbers properly?. Now I check my multiplication for each partial product. >
* alignment OK. checking 1st partial product: 5x7=35, 5 carry 3; 5x0=0, plus 3 is 3 – yes, that's right. I'll check another – 5x6=30, plus 4 = 34 – yes, that's right. checking 2nd partial product: 6x7=42, 2 carry 4; I'll check another -– 6x4=48, 8 carry 4, right *

Check addition (need not check all the addition links).
< Now I check some of my addition columns >
* 5+0=5; 3+2=5; Now I'll check another one – 4+8=12, 2 carry 1; 1+3+0=4, yes that's right." *

```
     4  4
     4  3
     6 8 0 7
   X     6 5
   3 4 0 3 5
 +4 0 8 4 2 0
  4 4 2 4 5 5
```

Step 2
Student

The same problem is given to the student on a clean sheet of paper.

The student solves the problem while verbalising the steps of the strategy just demonstrated with the help of prompts from the experimenter if necessary. Students were reinforced verbally if their verbalisation corresponded with their ongoing performance.

Step 3
Experimenter

Demonstrates the other checking procedure using the same problem on a clean sheet of paper.

"This is the procedure where we multi-check (check as we go)."

"We first multiply the multiplicand by the units digit of the multiplier. We multiply each digit of the multiplicand by the digit of the multiplier in one direction and double check by using the reverse direction.

< We first multiply 6807 by 5 ones and double-check in the reverse direction as we go >

* 5x7 is 35, 5, carry 3, (writes); 7x5 is 35, 5, carry 3; 5x0 is 0, plus 3 is 3, (writes); 0x5 is 0, plus 3 is 3; 5x8=40, 0 carry 4, (writes); 8x5=40, 0 carry 4; 5x6=30, plus 4, equals 34, (writes); 6x5=30, plus 4, equals 34 *

Now we multiply the multiplicand by the tens digit of the multiplier. So put a zero and shift numerals one place to the left, that is, indent one place to the left.
Double check – Zero is there.
We multiply each digit of the multiplicand by the digit of the multiplier in one direction and double check by using the reverse direction.

< Now we multiply 6807 by 6 tens so we put a zero and move one place to the left. As we multiply we double-check in the reverse direction >

* Place one zero. 6x7=42, 2 carry 4, (writes); 7x6=42, 2 carry 4; 6x0=0, plus 4 is 4, (writes) 0x6=0, plus 4 is 4; 6x8=48, 8 carry 4 (writes); 8x6=48, 8 carry 4; 6x6=36, plus 4 is 40, write 0 then 4; 6x6=36, plus 4 is 40 *

Check alignment, and Add.
Add in one direction and double check by adding in the reverse direction.
< Have I aligned numbers properly? – Yes. Now add >
* 5+0=5, (writes); 0+5=5; 3+2=5, (writes); 2+3=5; 0+4=4, (writes); 4+0=4; 4+8=12, 2 carry 1, (writes); 8+4=12, 2 carry 1; 3+0+1=4, (writes); 1+0+3=4; and 4+0=4, (writes); 0+4=4" *

Step 4
Student

The same problem is given to the student on a clean sheet of paper.

The student solves the problem while verbalising the steps using the strategy just demonstrated with the help of prompts from the experimenter if necessary. Students are reinforced verbally if their verbalisation corresponded with their ongoing performance.

Step 5
Student

Student completes worksheet independently (prompts provided if necessary) for 20 min. using either the mc, ec, or nc procedure as instructed by the experimenter.

Step 6
Experimenter

Feedback is given.

SAMPLE SCRIPT FOR THE DIVISION TRAINING SESSION

Step 1
Experimenter

Demonstrates the first checking procedure (e.g., multi-checking)
problem: 36 2 5 1 8 1

"This is the procedure where we multi-check (check as we go)."

"Write 5 and 10 times the divisor as a guideline.
< 36X5=180, 36X10=360 >

Determine the first partial dividend. Place dot.
< Which is the first biggest number that 36 will go into? >
* Does 36 go into 2? Does 36 go into 25? Does 36 go into 251? yes. Place dot above 1. *

Check – is partial dividend bigger than divisor?
< Check – Is 251 bigger than 36? Yes >

[a]Divide – Estimate largest multiple of the divisor.
< Get largest multiple of 36 which is less than 251 >.
* 251 is closer to 36x5, multiply by numbers greater than 5; 36x6=216, 36x7=252, can't use 36x7 as 252 is bigger than 251 so it is 36x6. *

[b]Place (initial) term of quotient above dividend, checking alignment. Check multiplication again in reverse direction.
< Place 6 above the dot. Check – multiply in reverse direction, that is, multiply 6x36. Check – Is it identical to 216? >
* I place 6 above the dot. OK. 6x6=36, 6 carry 3; 6x3=18, plus 3=21. Answer is 216. Yes it is identical. *

[c]Subtract multiple from dividend.
< 251-216=35 >

[d]Check subtraction by reverse operation of addition.
< Add 35 to 216 >
* 5+6=11, 3+1=4, 0+2=2 *

[e]Is subtraction answer greater than divisor? If yes, check working. If no, bring down next digit. Is this digit aligned properly?
< Is 35 greater than 36? No so bring down next digit which is 8. Check alignment. We get 358 as the nexy partial dividend. >

Repeat steps[a-e] until all digits of dividend are brought down.

Put up remainder if not zero."

```
        6 9 9 r 17
     ___________
  36|2 5 1 8 1
     -2 1 6
       3 5 8
      -3 2 4
         3 4 1
        -3 2 4
           1 7
```

Step 2
Student

The same problem on a clean sheet of paper is given to the student.

The student solves the problem while verbalising the steps of the strategy just demonstrated with the help of prompts from the experimenter if necessary. Students are reinforced verbally if their verbalisation correspondend with their ongoing performance.

Step 3
Experimenter

Demonstrates the other checking procedure using the same problem on a clean sheet of paper.
"This is the procedure where we check at the end."

"Write 5 and 10 times the divisor as a guideline.
< 36X5=180, 36X10=360 >

[a]Determine the first partial dividend. Place dot.
< Does 36 go into 2? Does 36 go into 25? Does 36 go into 251? yes. Place dot above 1. >

[b]Divide — Estimate largest multiple of the divisor.
< Get largest multiple of 36 which is less than 251 >.
* 251 is closer to 36x5, multiply by numbers greater than 5; 36x6=216, 36x7=252, can't use 36x7 as 252 is bigger than 251 so it is 36x6. *

[c]Place (initial) term of quotient above dividend, aligning it properly.
< Place 6 above the dot >

[d]Subtract multiple from dividend.
< 251-216=35 >

[e]Bring down next digit, aligning it properly.
< Bring down next digit which is 8. Check alignment. We get 358 as next partial dividend. >

Repeat steps[a-e] until all digits of dividend are brought down.

Put up remainder if not zero.

Now end-check:
Check alignment. Check if all digits have been brought down.
< Yes, all numbers are aligned and all the digits have been brought down >

Check each partial dividend. Check multiplication and subtraction (recheck at least half of the multiplication and subtraction links of each partial dividend)

< Check the first partial dividend >
* Yes the largest multiple of 36 is 6. Checking multiplication 36x6 = 216; checking subtraction – yes I get 35. It is smaller than 35. 8 is brought down to make 358 *

< Checking the next partial dividend. >
* O.K. 9X6=54, carry 5; 324 seems right; 358-324. 8-4=4, O.K.; 3-3=0; next digit down is 1 – O.K. *

< Checking the next partial dividend. >
* Next one is 341, check multiplication of 36x9 – 3x9=27, +5=32, seems right; check subtraction – 11-4=7, yes, it's right; 3-3= 0; 17 brought up. Yes, that's right" *

Step 4
Student

The same problem on a clean sheet of paper is given to the student.

The student solves the problem while verbalising the steps using the strategy just demonstrated with the help of prompts from the experimenter if necessary. Students are reinforced verbally if their verbalisation corresponded with their ongoing performance.

Step 5
Student

Student completes worksheet independently (prompts provided if necessary) for 20 min. using either the mc, ec, or nc procedure as instructed by the experimenter.

Step 6
Experimenter

Feedback is given.

Appendix 3

EACH STUDENT'S RAW DATA FOR EACH ASSESSMENT PERIOD ACROSS ALL EXPERIMENTAL PHASES

Table 12: Sue's accuracy data for all problems (O), single-digit multiplier problems (S), double-digit multiplier problems (D), and triple-digit multiplier problems (T), for each assessment period across all experimental phases

data points	baseline				verbalisation				alternating treatments: no-checking				alternating treatments: end-checking				alternating treatments: multi-checking				best treatment 1				follow-up				best treatment 2			
	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T
1	2/14	2/6	0/4	0/4	1/13	1/4	0/4	0/5	2/12	2/4	0/4	0/4	3/8	2/3	1/3	0/2	4/7	2/2	2/3	0/2	9/12	4/4	3/4	2/4	5/14	5/5	0/3	0/6	8/10	3/3	3/4	2/3
2	2/24	2/8	0/8	0/8	1/14	1/4	0/5	0/5	7/16	2/5	3/5	2/6	2/10	1/3	1/4	0/3	5/9	2/3	2/3	1/3	10/12	4/4	3/4	3/4	8/15	4/5	3/5	1/5	9/11	4/4	4/4	1/3
3	2/16	2/6	0/5	0/5	1/15	1/5	0/4	0/6	7/16	4/5	1/5	2/6	8/15	5/5	1/5	2/5	9/12	3/4	3/4	3/4	10/12	4/4	3/4	3/4	7/14	4/4	0/4	3/6	9/11	4/4	3/4	2/3
4	5/24	5/8	0/8	0/8	4/18	4/6	0/5	0/7	6/15	3/5	2/5	1/5	7/12	3/4	3/4	1/4	10/13	3/4	5/5	2/4												
5									10/15	5/5	4/5	1/5	9/12	4/4	4/4	1/4	10/12	4/4	4/4	2/4												
6									8/14	3/5	4/5	1/4	9/13	4/5	3/4	2/4	13/14	5/5	4/4	4/5												
7									8/15	4/5	3/5	1/5	9/13	3/4	3/5	3/4	10/12	4/4	3/4	3/4												

Table 13: Joe's accuracy data for all problems (O), single-digit multiplier problems (S), double-digit multiplier problems (D), and triple-digit multiplier problems (T), for each assessment period across all experimental phases

									alternating treatments																							
data points	baseline				verbalisation				no-checking				end-checking				multi-checking				best treatment 1				follow-up				best treatment 2			
	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T
1	2/11	2/5	0/3	0/3	1/8	1/2	0/2	0/4	4/12	2/4	1/4	1/4	1/8	1/3	0/3	0/2	4/5	1/1	1/2	2/2	10/12	4/4	3/4	3/4	8/16	5/5	1/5	2/6	10/11	3/3	4/4	3/4
2	1/12	1/4	0/4	0/4	1/11	1/3	0/4	0/4	9/16	4/5	3/5	2/6	3/9	2/3	0/3	1/3	7/9	2/3	3/3	2/3	10/12	4/4	3/4	3/4	12/15	5/5	4/5	3/5	13/14	5/5	5/5	3/4
3	1/9	1/4	0/3	0/2	1/11	1/3	0/4	0/4	8/14	3/4	3/5	2/5	9/16	3/6	4/5	2/5	9/11	2/4	4/4	3/3	12/14	4/5	4/5	4/4	12/16	5/5	3/5	4/6	14/14	5/5	5/5	4/4
4	0/7	0/2	0/2	0/3	2/14	2/5	0/5	0/4	11/16	5/5	4/5	2/6	8/13	3/4	4/5	1/4	9/12	4/4	3/4	2/4												
5									13/16	5/6	5/5	3/5	8/14	4/5	3/5	1/4	9/11	4/4	4/4	1/3												
6									10/15	5/5	3/5	2/5	9/13	4/5	2/4	3/4	9/12	3/4	3/4	3/4												
7									10/15	5/5	1/5	4/5	9/14	4/5	3/5	2/4	10/11	3/4	4/4	3/3												

Table 14: Vera's accuracy data for all problems (O), single-digit multiplier problems (S), double-digit multiplier problems (D), and triple-digit multiplier problems (T), for each assessment period across all experimental phases

									alternating treatments																							
data points	baseline				verbalisation				no-checking				end-checking				multi-checking				best treatment 1				follow-up				best treatment 2			
	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T
1	2/24	2/8	0/8	0/8	3/24	3/8	0/8	0/8	4/18	3/5	1/7	0/6	1/12	1/4	0/4	0/4	5/17	2/5	3/6	0/6	12/14	4/4	5/5	3/5	17/24	7/8	7/8	3/8	11/12	4/4	4/4	3/4
2	2/24	2/8	0/8	0/8	3/24	3/8	0/8	0/8	6/15	4/5	2/5	0/5	4/11	1/3	2/4	1/4	4/9	2/3	1/3	1/3	12/12	4/4	4/4	4/4	18/24	6/8	7/8	5/8	12/13	5/5	4/4	3/4
3	4/24	4/8	0/8	0/8	4/24	4/8	0/8	0/8	14/16	5/5	5/5	4/6	9/12	3/4	3/4	3/4	10/10	3/3	4/4	3/3	11/12	3/4	4/4	4/4	18/24	6/8	5/8	7/8	13/14	5/5	4/5	4/4
4	6/24	6/8	0/8	0/8	5/24	5/8	0/8	0/8	15/20	5/6	4/6	6/8	8/15	4/5	2/5	2/5	10/12	4/4	3/4	3/4												
5									15/17	6/6	5/6	4/5	12/14	5/5	4/5	3/4	11/12	4/4	4/4	3/4												
6									12/16	4/5	5/5	3/6	11/14	4/5	5/5	2/4	10/12	4/4	3/4	3/4												
7									14/16	5/5	4/5	5/6	12/14	3/5	5/5	4/4	11/12	3/4	4/4	4/4												

Table 15: Keith's accuracy data for all problems (O), single-digit multiplier problems (S), double-digit multiplier problems (D), and triple-digit multiplier problems (T), for each assessment period across all experimental phases

data points	baseline				verbalisation				alternating treatments: no-checking				alternating treatments: end-checking				alternating treatments: multi-checking				best treatment 1				follow-up				best treatment 2			
	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T
1	2/15	2/5	0/5	0/5	4/17	4/5	0/6	0/6	12/18	4/5	5/7	3/6	4/9	3/3	0/3	1/3	6/10	2/3	3/4	1/3	14/15	5/5	4/5	5/5	18/24	8/8	7/8	3/8	20/23	7/7	6/8	7/8
2	2/16	2/4	0/6	0/6	5/17	5/5	0/5	0/7	9/18	5/6	3/6	1/6	10/14	5/5	4/4	1/5	6/9	2/3	3/3	1/3	13/15	5/5	5/5	3/5	19/22	7/7	7/8	5/7	14/18	5/6	6/6	3/6
3	4/17	4/6	0/6	0/5	5/19	5/6	0/6	0/7	11/19	5/6	3/6	3/7	14/18	6/6	5/6	3/6	10/15	5/5	4/5	1/5	14/15	5/5	4/5	5/5	18/24	6/8	8/8	4/8	17/18	6/6	6/6	5/6
4	5/17	5/5	0/5	0/7	3/18	3/6	0/5	0/7	14/19	6/6	3/6	5/7	14/16	5/5	5/5	4/6	14/16	5/5	5/5	4/6												
5									14/18	5/6	6/7	3/5	12/15	5/5	4/6	3/4	9/12	3/4	3/4	3/4												
6									16/18	5/5	7/7	4/6	15/15	5/5	5/5	5/5	13/14	5/5	3/4	5/5												
7									14/18	6/6	5/5	3/7	15/16	6/6	5/5	4/5	12/14	5/5	4/5	3/4												

Table 16: Mary's accuracy data for all problems (O), single-digit multiplier problems (S), double-digit multiplier problems (D), and triple-digit multiplier problems (T), for each assessment period across all experimental phases

									alternating treatments															
data points	baseline				verbalisation				no-checking				end-checking				multi-checking				best treatment			
	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T
1	3/17	2/6	1/6	0/5	5/20	3/6	1/7	1/7	10/18	2/5	5/7	3/6	4/6	1/2	1/2	2/2	5/11	2/3	1/4	2/4	12/16	4/5	4/5	4/6
2	5/18	3/5	1/6	1/7	2/19	1/5	1/7	0/7	11/17	4/6	4/5	3/6	11/15	4/5	3/4	4/6	9/12	3/4	3/4	3/4	13/16	6/6	4/5	3/5
3	5/19	4/7	1/6	0/6	3/19	2/6	1/6	0/7	13/21	5/7	4/7	4/7	12/18	4/6	4/6	4/6	11/14	3/5	5/5	3/4	13/16	4/5	6/6	3/5
4	5/21	4/6	0/7	1/8	3/18	3/6	0/5	0/7	15/20	5/6	6/6	4/8	13/17	5/5	4/6	4/6	12/15	4/5	5/5	3/5				
5									9/19	3/6	4/7	2/6	11/16	5/5	4/6	2/5	12/15	3/3	6/7	3/5				
6									15/19	5/5	6/7	4/7	11/14	4/5	5/5	2/4	12/15	5/5	4/5	3/5				

Table 17: Jane's accuracy data for all problems (O), single-digit divisor problems (S), double-digit divisor problems (D), and triple-digit divisor problems (T), for each assessment period across all experimental phases

									alternating treatments																							
data points	baseline				verbalisation				no-checking				end-checking				multi-checking				best treatment 1				follow-up				best treatment 2			
	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T
1	3/24	3/8	0/8	0/8	3/24	3/8	0/8	0/8	0/5	0/2	0/2	0/1	1/5	0/2	0/2	1/1	2/5	1/2	1/2	0/1	8/9	3/3	3/3	2/3	9/12	3/4	2/4	4/4	8/9	3/3	3/3	2/3
2	3/24	3/8	0/8	0/8	4/24	4/8	0/8	0/8	2/8	1/2	0/3	1/3	4/7	2/3	1/2	1/2	6/7	2/2	2/3	2/2	8/8	3/3	3/3	2/2	9/11	4/4	4/4	1/3	10/10	3/3	4/4	3/3
3	4/24	4/8	0/8	0/8	2/24	2/8	0/8	0/8	7/10	2/3	2/4	3/3	6/7	2/2	3/3	1/2	7/8	3/3	3/3	1/2	9/9	3/3	3/3	3/3	9/11	3/4	4/4	2/3	10/10	3/3	3/3	4/4
4	2/24	2/8	0/8	0/8	3/24	3/8	0/8	0/8	7/11	3/4	3/4	1/3	6/8	2/3	2/3	2/2	8/9	3/3	3/3	2/3												
5									7/11	2/4	3/4	2/3	6/9	1/3	2/3	3/3	7/8	3/3	3/3	1/2												
6									8/10	4/4	2/3	2/3	7/8	2/2	2/3	3/3	8/8	3/3	3/3	2/2												

Table 18: Joan's accuracy data for all problems (O), single-digit divisor problems (S), double-digit divisor problems (D), and triple-digit divisor problems (T), for each assessment period across all experimental phases

data points	baseline				verbalisation				alternating treatments: no-checking				alternating treatments: end-checking				alternating treatments: multi-checking				best treatment 1				follow-up				best treatment 2			
	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T	O	S	D	T
1	2/24	2/8	0/8	0/8	4/24	4/8	0/8	0/8	2/6	0/2	1/3	1/1	1/3	0/1	1/1	0/1	2/4	1/2	1/1	0/1	10/10	4/4	3/3	3/3	12/14	4/4	4/6	4/4	13/13	4/4	4/4	5/5
2	3/24	3/8	0/8	0/8	4/24	4/8	0/8	0/8	2/8	1/2	1/3	0/3	4/8	2/3	1/3	1/2	6/7	2/2	2/3	2/2	10/11	4/4	4/4	2/3	11/13	4/4	4/5	3/4	13/13	5/5	4/4	4/4
3	4/24	4/8	0/8	0/8	4/24	4/8	0/8	0/8	6/8	2/2	2/3	2/3	5/6	2/2	2/2	1/2	8/8	3/3	3/3	2/2	10/10	4/4	3/3	3/3	11/13	3/4	5/5	3/4	12/12	4/4	4/4	4/4
4	3/24	3/8	0/8	0/8	4/24	4/8	0/8	0/8	9/10	3/4	3/3	3/3	6/8	2/3	2/3	2/2	11/11	4/4	4/4	3/3												
5									13/13	4/4	5/5	4/4	7/8	2/2	2/3	3/3	7/8	3/3	3/3	1/2												
6									12/12	4/4	4/4	4/4	9/9	3/3	3/3	3/3	12/12	4/4	4/4	4/4												
7									9/11	4/4	3/4	2/3	7/9	2/3	2/3	3/3	9/9	3/3	3/3	3/3												