ASSESSMENT, SELF-ASSESSMENT AND PERFORMANCE

----------------------------------------

ANTHONY J. PICKUP

Acknowledgements

Thanks are due to the head teachers and participating teachers of the main study, and the student assistant experimenters of the laboratory study.

Thanks are also due to Dr. B. Baughan and Dr. Judith Bailey, University of London, for assistance in setting-up computing, and to Professor W. James, Dean of the Faculty of Educational Studies, the Open University, for provision of study-leave and secretarial facilities, for the writing of this report.

As for my M.Ed thesis of 1967, my special thanks go to my supervisor, Dr. W.S. Anthony, for his insight, assistance and constant encouragement over the five years of part-time work which led to this report.

Anthony J. Pickup
CONTENTS

1. Introduction. Page -1-

PART I

The Laboratory Study

2. Background to the Laboratory Study. -5-

3. The Laboratory Experiment: Method. -20-

4. The Results of the Laboratory Experiment. -36-

5. The Laboratory Study: Conclusions. -59-

PART II

The Main Study

6. The Background to the Main Study. -62-

7. The Main Study: Method. -69-

8. The Main Study: Results. -80-

9. The Main Study Results: Discussion. -92-

PART III

10. Conclusions and Implications. -104-

Appendices. -117-

References.
CHAPTER 1
INTRODUCTION

This thesis presents the report of two experimental investigations of the motivational function of feedback upon performance in educational contexts. The work was commenced in 1968, and was derived from a previous set of investigations carried out in part fulfilment of the conditions for the degree of M.Ed. This previous work was further refined, and these refinements and the major findings were published in the British Journal of Educational Psychology. A copy of this publication is bound in this volume as an Appendix. Because the present study is derived so closely from the author's previous work, the traditional 'Review of Literature' is omitted, and replaced by reviews of similar research at appropriate points in the text, particularly in those chapters concerned with conclusions drawn from the present research.

The thesis is in three parts. Part I presents a report of laboratory-type investigation of the topic. Part II describes research conducted in classroom settings. Part III is concerned with the major conclusions of both of these investigations, and attempts to relate the present study to other published research, where this is possible. Parts I and II are similarly composed, each having a chapter which introduces the rationale of the study, a chapter describing the method, a chapter presenting the findings (without discussion) and a chapter which discusses the findings. Part III is a single chapter of discussion, relating major findings to other similar studies.

While detailed description of the rationale behind each study is presented in Parts I and II, the starting-point for this research is illustrated by the following quotation from
Pickup and Anthony (1968)

"From the present study it seems that if there is a single motivational effect of a discrepancy between a pupil's expected mark and the mark received from the teacher, this effect is so small as to be insignificant in the normal classroom setting. Such an effect might however appear with the use of more sophisticated performance tests and laboratory-type conditions. Other experiments employing a method similar to that employed in the main study could investigate the effects of repeated discrepancies over a larger period of time".

The research carried out prior to the present study (subsequently to be described as "previous work") investigated pupils' expected marks for a piece of work and their relationship to the teacher's mark for that work. The effects of discrepancies between expected marks and teachers marks upon a subsequent performance constituted a major aim of the experimental investigations. No single motivational effect could be demonstrated for either positive or negative discrepancies; though negative discrepancies appeared to be more effective incentives to pupils in the upper ranges of ability, and positive discrepancies to be more effective in the bottom ranges of ability. It was considered that this effect (termed the 'reversal effect') could operate to obscure a statistically significant effect of treatment, if the positive effects of treatment at one level of ability is cancelled-out" at another level.

The previous work contained no procedures by which the
artificial distortion of marks for motivational purposes could be evaluated, as an incentive strategy, against a strategy where no artificial distortion of marks took place. The present study incorporates these procedures.

In the period since the previous work was published it has become clear that the term 'expectation' is now more frequently associated with a specific form of 'aspiration' phenomena. Accordingly, this term and its derivatives (eg. "expected mark") will be replaced by the term 'self-assessment', except where reference to previous work requires the use of the original terminology.
PART I

THE LABORATORY STUDY
INTRODUCTION

Previous experimental investigation of the short-term effects of discrepancies between marks expected and marks received (Pickup 1967, Pickup and Anthony 1968) suggested further investigations. It had been shown that, if a single effect could be ascribed to a particular type of discrepancy this effect was so small as to be insignificant in normal classroom contexts. A further and unanticipated finding was that the incentive effect of one treatment relative to its opposite reversed as the opposing treatment were applied at differing levels of pre-test score. This latter finding (significant at the .02 level) was a possible explanation for the former insignificant results.

It would appear that a particular treatment could have a positive incentive effect at one level of ability and a neutral, or even negative effect at other levels of ability. The evidence for this interpretation was however, subject to considerable contamination.

The experimental design from which the results has been derived was one in which the aim of generalisability to normal classrooms had been a priority. This priority, and the need to avoid any 'Hawthorne Effect' had necessitated the use of tasks
and measures which were normal features of the
classrooms in which the study took place. This
type of investigation had indicated that further
development and investigation in classroom
contexts was both feasible and desirable. Part II
of the present study is concerned with this
development and investigation.

There was also evidence for the existence of
hitherto uninvestigated factors in the motivation
of learning, which could have implications for the
field of 'Knowledge of Results'. Investigations in
this area were seen to require more closely
controlled experimentation than was possible in
classroom contexts. This part of the present study
is concerned with research conducted in a laboratory
context; designed to explore the findings of the
previous work with improved experimental rigour.

THE BACKGROUND TO THE LABORATORY STUDY

The previous work had demonstrated that there
was no significant difference in classroom
performance between groups exposed, on a single
occasion, to one of the three experimental treatments
of:
a) Receiving a mark greater than expected
('positive' treatment),
b) Receiving a mark less than expected ('negative' treatment) and
c) Receiving a mark equal to that expected ('matching' treatment).

Further findings indicated that these treatments appeared to have some effects contingent upon uncontrolled factors. The presence of these effects lent some support to the theoretical base of the previous work and indicated that a new factor was 'intervening' between treatment and performance.

The reasons that previous work had failed to demonstrate a single incentive effect for a particular experimental treatment could therefore be:
1. The presence of an uncontrolled variable intervening between independent variable and dependent variable.
2. Other sources of experimental error.

An examination of these areas was conducted with the purpose of achieving an improved design.

1. The uncontrolled variable

The previous work had shown that personality variables and academic motivation were of less value as predictors of response to a particular treatment than was level of pre-test score. This finding could indicate that treatment effect was related to ability. However, the phenomenon occurred in classes
of differing academic ability (in streamed school contexts) and a more likely explanation was that self-perceived ability (or academic status) relative to classmates was a factor which could modify the effect of a particular experimental treatment.

If this were the case it could follow that the effect of an experimental treatment would depend in part upon the climate and history of a particular classroom context, and in part upon the individual within this context. The precision with which an individual could place himself relative to his classmates would depend upon the element of competition, or the degree of publicity of marks and grades within the classroom context. There would also be considerable variation both within and between individuals in the importance with which such information would be viewed.

The variation of effect of a single experimental treatment indicated that the return of a mark was likely to carry two types of motivational information; 'individual'; related to the individuals private assessment of his work, and 'social' where the individual was conscious of status relative to classmates. Additionally, the salience of the social component of an experimental treatment
could be hypothesized as being greater at the extremes of ability, since the crude classifications of being at the 'top' or the 'bottom' of a class are clear status signposts.

The particular effects noted in the previous work are consistent with the hypothesis that each experimental treatment contained an 'individual' and a 'social' component, and that the incentive effect of the social component was activated in situations where the individual was concerned to preserve or improve upon a previously perceived status position.

Other sources of experimental error

Since a major aim of the previous work was that the findings be generalisable to normal school contexts, a wide range of common school tasks were included in the investigation. The tasks which provided the criterion measures for the pre-tests and post-tests of performance were thus a sampling of those school activities which were amenable to measurement by objective-type tests. Pupils' performance was tested on tasks as diverse as spelling ability, arithmetic, modern mathematics and English comprehension. Each class was measured twice in the same subject area within a short period of time, but the content of the pre and post-tests could be less than satisfactorily parallel in the
range and type of skills or knowledge required.

The criterion measures were teacher-constructed tests of normal classwork prepared to a simple specification provided by the experimenter (Pickup 1967, P.72) to avoid serious divergence from a normal distribution of scores. The reliability indices for these tests are in appendix A, and indicate that while they are more reliable than might have been expected, they nevertheless fall short of the reliability which is desirable in a rigorous investigation of complex incentive action.

Similarly, the employment of the usual classroom teacher in the administrative phases of the experiment, while avoiding any bias from either the 'Hawthorne effect' or from 'experimenter effect' introduced the possibility of considerable variation in the experimental conditions. This was an acceptable feature of the generalisable design (there would be such variation in normal classroom contexts, after all), but was clearly unacceptable in a more rigorous investigation.

THE LABORATORY STUDY: RATIONALE AND DISCUSSION
This study (as the previous studies) had the general aim of investigating the phenomena which existed in a classroom when, following the performance of a task, a mark was returned to a pupil by a teacher. The situation may be described in more fundamental terms as one in which an assessment of task performance is returned to a performer by some agency external to himself. This and the previous studies in this area make a primary assumption that a performer has some cognition regarding his own performance, and that this cognition may be observed by requesting the individual to state the score which he believes his work to merit. This measure is termed the subject's expectation.

The relationship of expectation and mark returned is a situation which contains potential to motivate improved subsequent performance. The experiments in this area are designed to:

(i) investigate this potential, and
(ii) evaluate the strategy of artificially distorting marks for motivational purposes.

The objective of the laboratory study was to replicate and extend the field of investigation of previous work in conditions of improved experimental control, and was designed as a progression from the
previous work in two major respects:

(a) A reduction of general experimental error.
(b) The extension of experimental control.

More specifically, the laboratory study was designed to

(1) Reduce the error arising from the variety of conditions under which previous work had been carried out by standardising the conditions for all experimental groups.

(2) Reduce the error associated with the measurement of performance by (a) ensuring that all tasks were of equivalent difficulty. (b) standardising and simplifying the nature of the tasks, and (c) improving the identification of a correct response.

(3) To investigate a potentially interacting 'social' variable which could have contributed to the "reversal effect" noted in Pickup and Anthony (1967).

(4) To extend investigation to include observation of undistorted feedback under the same experimental conditions as investigation of distorted or manipulated feedback.
The research questions

1. Do any of the experimental treatments of earlier work (subsequently described as 'individual' treatments) prove significant incentives in a laboratory context where experimental error is reduced?

2. Does information of a social 'status' nature have an incentive effect?

3. Does 'social' information interact with 'individual' information to either reduce or enhance any incentive effect, and is any combination of social and individual information an incentive for improved performance?

4. Are the findings of earlier work with regard to the phenomena of expected score confirmed or disconfirmed in a laboratory context and further, what are the relationships between performance, expected score and experimental treatments?

5. Are the incentive effects of manipulating feedback (artificial distortion of marks) greater than might occur with the return of actual scores?
The phenomena under study:

1. Performance of subjects in the type of task commonly found in educational contexts.
2. The individual's expected mark (or self assessment) for this performance.
3. The type of feedback which the individual receives from a source external to himself.

The above phenomena may be represented as variables within the design of an experiment as follows:

Dependent variable (a) Performance.

The observation of performance change following experimental treatment was the major aim of the laboratory experiment. The limitations of a laboratory study for generalisation to a school context are obvious, but generalisability could be improved by the selection of a task which was closely related to common school tasks particularly in the use of attention and memory. A simple cancellation task would not be appropriate. It was, however, recognised that any laboratory task could be no more than an approximation to normal school tasks.

The observation of performance over a series of trials implies the use of task versions of equivalent
difficulty. Efficiency of data collection and the retention of some elements of performance in a social setting dictated that the task be capable of group administration.

A verbal learning task was selected as the performance criterion. Subjects were asked to recall nonsense syllables which had been projected onto a viewing screen. The task is described in detail in Chapter 3.

**Dependent variable** (b) **Expected Score**.

The expected score was the basis from which experimental treatments were derived (see below) but could also be observed as a dependent variable in its own right. In particular, the observation of the 'accuracy' with which an individual may perceive the quality of his performance (see SAI below) was of interest.

**The independent variable: Type of feedback.**

The laboratory study was designed to include a systematic standardised representation of conditions which could occur in normal educational feedback situations, together with a comparison group receiving 'variable' (actual score) feedback. The experimental treatments therefore included the following:
(a) Receiving a mark greater than expected.
(b) Receiving a mark less than expected.
(c) Receiving a mark equal to that expected.
(d) Receiving an actual score

And, in combination with the above, Either

(e) Being given no information about performance relative to other individuals,
Or,

(f) Being given information that performance was superior to that of other individuals,
Or,

(g) Being given information that performance was inferior to that of other individuals.

The size of discrepancy selected for treatments (a) and (b) above was 10% of the range of scores possible on the task. It was considered that this discrepancy was large enough to produce an incentive effect, but not so large as to be an obvious distortion of true performance.

The simple datum of an 'average' performance (not specified) was used in treatments (f) and (g) above. Additionally, an experimental group which received no feedback of any kind was included in the experiment.
In view of the small discrepancy size it was considered that significant differences in performance might occur only after two applications of an experimental treatment. The novelty of the procedures to the experimental subjects was an additional factor which required consideration. It was decided that the design should include an adaptation trial followed by a pretest, and two post-tests.

General discussion.

The transfer of real-life social phenomena to laboratory settings involves a sacrifice of realism in exchange for control. In this case the advantages of conducting a laboratory investigation went further than simple reduction of measurement error. The nature of the social variable, and the laboratory context of this information provide a structure where qualms about the ethics of providing distorted information (which may be damaging to the subjects self-concept) are reduced. Few subjects would be distressed to learn that they were 'below average' in the learning of nonsense syllables. The experimental investigation of this type of information in classroom contexts would be fraught with both technical and ethical difficulties. It was recognised, however, that the effects of the laboratory variable
upon performance would be similarly smaller than effects likely to occur in real-life situations.

There was considered to be no severe ethical problem where 'individual' treatments were concerned, as previous studies had shown that discrepancies between expectation and true score considerably larger than 10% occurred in real-life situations.

The Self-Assessment Index (S.A.I.)

The laboratory experiment, the main study (Part II) and the previous work has been designed to investigate the influence upon performance which may result from various forms of feedback, in relation to self-perception of performance. This general model also permits the study of the relationship between (i) subjective self-assessment of performance and (ii) assessment of performance derived from sources 'outside' the individual; in other words, a comparison between 'interior' (to the individual) and 'exterior' forms of assessment.

The discrepancy between an 'exterior' assessment (such as performance scores in a laboratory setting) and the 'interior' assessment (expected score) is a phenomenon which is the subject of some of the observations reported in both the laboratory experiment and in the main study (Part II). It is
proposed to coin the term "Self-Assessment Index" (S.A.I) as a description of the discrepancy between the value of an 'exterior' assessment and the value of an 'interior' assessment.

In the laboratory study the S.A.I. represents the numerical difference between a performance score and an 'expected', or self-assessed score from that individual. In the main study (part II) the S.A.I represents the numerical difference between a teachers mark and the pupils' self-assessment for a classroom test. (S.A.I.=Self-assessment minus exterior assessment).

In both studies the S.A.I is a variable which has a theoretical range double that of the performance measure on which it is based and can exist in a positive or negative form, with a zero midpoint.

**Additive effect of treatment**

For both the laboratory study and the main study the assumption is made; that the effects of successive applications of treatment will proceed in the same direction as earlier effects of that treatment: and that successive applications of treatment are at least additive in their effect.
CHAPTER 3

THE LABORATORY EXPERIMENT: METHOD

Introduction

The purpose of this experiment was to examine the influence of two repetitions of 13 different forms of feedback upon the performance of individuals in verbal learning tasks. The experimental treatments consisted of differing types of 'feedback' or 'knowledge of results' selected to represent combinations of two factors of feedback.

1. 'Individual' treatments, concerned with the performance and expectation of the individual, and

2. 'Social' treatments, concerned with the performance of that individual relative to that of other individuals.

The experiment was based on the presentation of four tasks of equivalent difficulty, presented in four separate experimental trials. The first trial was an adaptation trial designed to familiarise all subjects to the experimental conditions, and for which the subjects received no feedback. At the conclusion of Trial II the first of two repetitions of experimental treatment was applied. The effects of this treatment were observed on the third trial, at the conclusion of which the second experimental treatment was applied. The effects of this
treatment were observed in Trial IV, and feedback of a non-experimental nature was supplied to the subjects to conclude the experiment.

The experiment was conducted in a fifteen-booth language laboratory. Fig.3.1 shows the layout of the laboratory.

Fig.3.1. Language Laboratory Plan

The experimental subject worked in a booth isolated from other subjects in the same sitting, and was asked to learn nonsense syllables which were displayed on a screen at the front of the room. By means of headphones installed in the booths, standard background noise - (quiet music) was played to all subjects during the time they were asked to recall the nonsense syllables which had been presented. Experimental instructions and feedback were also communicated by headphone.

Each trial consisted of the subject attempting to learn a set of ten nonsense syllables displayed on a screen for 20 seconds and then record what could be recalled on the special form provided in a period of one minute.
(See Appendix C for example of form). In addition to this task the subject was asked to record the number of nonsense syllables which he/she believed had been correctly recalled. This score represented the individual's 'expectation'.

At the conclusion of each trial the record form for each subject was collected and scored by an experimental assistant who supplied information to the experimenter which was then returned to the subject.

At the conclusion of the final trial the subjects true performance score was returned as an aid to disguising the nature of the experiment from future groups of subjects.

Additionally, subjects were asked to complete a performance summary form (see Appendix B). The purposes of this document were:

1. To ensure that the subject had heard and understood the experimental feedback.
2. To provide a visual record of performance and expected score for those trials where feedback was supplied.

The performance measures

Four sets of 10, 3-letter, nonsense syllables were employed as tasks in the experiment. The nonsense syllables were graded in association value, (Hilgard, R, 'Methods and Procedures in the Study of Learning') so that each set of 10 syllables contained:
1) 2 syllables of 100% association value,
2) 2 syllables of 0% association value,
3) 6 syllables of intermediate values,
   between the ranges of 7% - 80% association value.

(See instructions to Scorers' Appendix D for complete
listing of nonsense syllables for all trials).

A description of one set is as follows:

<table>
<thead>
<tr>
<th>Association Value</th>
<th>No. of Syllables</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>2</td>
<td>LON, NAR</td>
</tr>
<tr>
<td>80%</td>
<td>1</td>
<td>KOM</td>
</tr>
<tr>
<td>73%</td>
<td>1</td>
<td>PIF</td>
</tr>
<tr>
<td>47%</td>
<td>1</td>
<td>MUB</td>
</tr>
<tr>
<td>27%</td>
<td>1</td>
<td>RIY</td>
</tr>
<tr>
<td>13%</td>
<td>1</td>
<td>GIK</td>
</tr>
<tr>
<td>7%</td>
<td>1</td>
<td>QEM</td>
</tr>
<tr>
<td>0%</td>
<td>2</td>
<td>YUD, YIL</td>
</tr>
</tbody>
</table>

Table 3.1: Association Values, number of syllables and Items.

The nonsense syllables for each set were displayed
in two groups of five, in a random sequence (Fig.3.2).
A permanent transparency slide was made of each of the four sets of syllables to permit the use of an overhead projector in the laboratory situation.

THE EXPERIMENTAL TREATMENTS

Twelve of the 13 experimental treatments were combinations of two factors:

1. Individual Treatments. 4 'levels' of this factor were employed. Using the expected score and the performance score produced by each individual as criteria, the following individual treatments were used.

   Individual Treatment 1. "Actual"
   Here, the subject was informed of his/her performance score in trials II and III.

   Individual Treatment 2. "Matching"
   Here, the subject was informed that his/her
performance score exactly matched his/her
'expected score' regardless of his/her true
performance.

Individual Treatment 3. "Negative"
Here, the subject was informed that his/her
performance score was one point less than his
expected score, regardless of his/her true
performance.

Individual Treatment 4. "Positive".
Here, the subject was informed that his/her
performance score was one point more than his
expected score, regardless of his/her true
performance.

2. Social Treatments. In combination with the
Individual Treatments, 3 'levels' of this factor
were used.

Social Treatment 1. "Non-Social"
Here, no information was given regarding the
performance of other individuals on this task.

Treatment 2, "Social negative"
Here, information was given that the subjects
performance (however good or bad it might be)
was "below average".
Treatment 3. "Social Positive"

Here, the individual was informed that his/her performance (however good or bad it may be) was an "above average" performance.

A Thirteenth Treatment named 'Control' was employed, where the individuals were not given any information with regard to their performance on any Trial.

No Treatment

If any individual in the 'Matching', 'Negative' or 'Positive' groups produced an expected score which would lead to feedback of 0 marks or 10 marks (or beyond these extremes) scorers were instructed to abandon experimental procedures for that individual and to assign that individual to the 'No Treatment' category. The subject remained present in the experimental session, but was returned his/her true score and the data from that individual was scrapped. 5 such cases occurred, and the individuals were replaced in subsequent sessions so that each experimental treatment group consisted of ten individuals.

Other variables

Other potentially intervening variables were controlled by standardisation and randomisation. The standardised procedures were

1. Time allowed for the task, equal for all trials and all subjects.
2. Equal distracting sound level (background music) while recalling nonsense syllables.

3. The use of a standard record form.

The randomised procedures were:


5. Allocation of experimental treatment to a particular subject/assistant experimenter pair.

The above was intended to ensure both random allocation of treatment to subjects, and to reduce experimenter/subject interaction by random allocation of treatment to assistant experimenter.

The Sample

The sample of 130 subjects involved in this experiment was arrived at as follows:

All male and female students in their second year at a London College of Education (excluding students who had volunteered to assist in the experimental procedures) were invited to participate in an experiment concerning factors in learning. The students were approached by the experimenter during a timetabled 'education' session and were informed that, while the experiment would be conducted in their lecture time, it was not part of their course and attendance was not compulsory. It was further mentioned that the experience of participation should be of use to them as students of learning. The results were to be treated as confidential.
Opportunity to decline participation was given to all students, but in the event, only one male student expressed unwillingness to participate. The students were then given appointments to attend experimental sessions.

Experimental sessions were normally conducted with 13 at a sitting. The loss of 8 subjects due to violation of experimental conditions and non-attendance (see under 'no treatment' category above) was made good by subsequent sessions of up to 15 subjects until each of the 13 experimental treatments had achieved a group size of 10 subjects.

The experimental assistants

The experimenter, in his role as lecturer to a group of 26 students, outlined the proposed experiment and asked for volunteers to assist in the conducting of a number of experimental sessions. 18 students volunteered to act as experimental assistants. These volunteers subsequently participated as subjects of a pilot experiment, where most of the experimental procedures were employed. Following this pilot experiment, volunteers then received a detailed account of the procedures to be employed (see Instructions: Scorers Appendix D) and were informed of the purposes of the research and of the required conditions. A further pilot session with the remaining 8 students in the education division was conducted, using the full experimental conditions.

Conduct of experimental sessions

The assistant experimenters reported to the laboratory
ten minutes before the scheduled session time and collected one 'Instructions to Scorer' (see Appendix D) leaflet at random from thirteen possible leaflets. These leaflets represented the possible experimental treatments. The assistant experimenter then chose a card at random from a box containing fifteen numbered cards (representing the fifteen booths available in the laboratory) to determine the booth for which that scorer would have responsibility in this experimental session. Thus, scorer/treatment and experimental treatment/booth position were randomised. Each assistant experimenter then placed 4 copies of a score sheet and 1 copy of a 'Performance Summary' (see Appendix B) in the booth for which he or she was responsible. Each scorer then retired to the scorers waiting area for the row in which his or her booth was situated.

The experimental subjects were brought from a waiting area to the laboratory as a group, and were invited to sit in any booth which had scoring forms on the desk. All subjects were then requested to put on the headphones for their booths. When all headphones were in place a pre-recorded tape of instructions to subjects was played (see Appendix E for transcript).

At the conclusion of the playing of the instruction tape the experimenter activated the microphone on the experimenters console so that each subject in each booth could hear subsequent instructions.
These were;

**Trial I**

Please insert 'Trial One' on your score sheet, if you have not already done so, and then place your pens or pencils on the table.

**Pause**

Trial one is about to start. 5-4-3-2-1"
The first set of nonsense syllables were then thrown onto the screen behind the experimenter for a period of 20 seconds. Then the projector was turned off. At this point quiet background music was fed to each experimental subject by the headphones in his booth. After a period of one minute of such background music the experimenter again activated the microphone so that his voice was heard over the background music, and said "Please remember to insert your expected score on both your score sheet and your Performance Summary. When you have done this, raise your hand to inform your scorer."

At the expiry of the two minute period, or when all booths had indicated to their scorers that they had completed the experimental procedures for this trial the experimenter again activated the microphone to say "Thank you, you all appear to have understood the experimental conditions. In a moment, we will commence the next trial."
Trial II

The experimenter, having selected and set up the transparency containing the material for the second trial of the experiment, activated the microphone in the experimenter console, and, speaking to all the subjects said, "Please insert trial number two and your booth number on your score sheet, and place your pens or pencils down."

Then after a short pause,

"Trial two is about to start. 5-4-3-2-1"

At this point the transparency for trial II was exposed on the screen behind the experimenters console for a period of 20 seconds. Again, when the projector was turned off, background music was played to all consoles. After a period of one minute of background music the experimenter activated the microphone to all subjects and said "When you have completed the details on both forms, raise your hand to inform your scorer."

As each subject completed both forms he or she raised his or her hand to inform the scorer. Then the scorer went to that booth to ensure that the forms had been completed and then removed them to administer the experimental procedures as outlined in "Instructions To Scorers." When the scorers had completed their procedures they brought the completed score sheet to the experimenter whom (by means of the selective switching function available on the console) selected a particular booth and informed the occupant of the booth as follows;
"booth X, your actual score is Y" (or; booth X your actual score is Y, and this is an above/below average performance," if that subject fell into one of the "social" conditions). This was followed by an instruction to enter this score on the performance summary for trial II. After each experimental subject had been informed of a score provided with the experimental feedback, trial III began.

**Trial III**

The experimenter activated the microphone to all subjects and said "Insert trial III on your score sheet, and please place your pens and pencils down." A short pause followed, then "Trial III is about to start. 5-4-3-2-1."

The transparency for trial III was exposed for 20 seconds followed by background music, followed in turn after a one minute pause by the experimenter activating the microphone to all subjects and saying "When you have completed details of both forms, please raise your hand." As the scorers came to the experimenters console each experimental subject was informed of his experimental feedback individually, as in trial II. When all experimental subjects had been provided with their experimental feedback this concluded trial III.

**Trial IV**

When all experimental subjects had completed Trial III and when they had been informed of their experimental feedback, the experimenter activated
the microphone to all subjects and said "Insert
Trial IV on your score sheet. Please put down pens
and pencils," short pause "Trial IV is about to
start. 5-4-3-2-1."

The transparency for trial IV was exposed on
the screen for 20 seconds and following this,
background music was played to the experimental
subjects. After one minute of background music the
experimenter activated the microphone to all subjects
and said "When you have completed details on both
forms, please raise your hand."

As the experimental subjects finished they
informed their scorers, who again collected the forms,
carried out the experimental treatments and brought
the score sheets to the experimenter who contacted
each booth individually to provide each subject
with his/her true score. When each individual
had been informed of his/her true score in the
same way as in trials II and III the microphone was
activated to the whole experimental group and
the experimenter said "Please insert group (here a
code designating the order in which experimental
sessions had been carried out was entered) on your
performance summary, and when you have done so please
raise your hand to inform your scorer." As the
subjects did this, the scorers collected the
performance summaries.

When all performance summaries had been
collected the experimenter again activated the
microphone to all subjects and said "This concludes the experiment. Please do not discuss this experiment or compare performances with anyone. I will be seeing all education divisions involved with a full report on the experiment as soon as possible. You will then see the need for secrecy! Thank you for your cooperation."

Null Hypotheses and Data analysis

Null hypotheses:

1. There will be no significant difference in performance on verbal learning tasks between groups exposed to two applications of thirteen experimental treatment combinations.

2. There will be no significant difference in performance gain on verbal learning tasks between groups exposed to two applications of four individual treatments.

3. There will be no significant difference in performance gain on verbal learning tasks between groups exposed to two applications of three 'social' treatments.

4. There will be no significant interaction between social and individual treatments as measured by performance gain on verbal learning tasks following two applications of treatment.
Possible methods of Analysis were;

Hypothesis 1 above tested by Analysis of Variance; or analysis of Covariance if pre-test performance scores were not equivalent.

Hypotheses 2, 3, and 4 tested by two-way analysis of variance.
Chapter 4

The Results of the Laboratory Experiment

SECTION A. Summary statistics

1. Distribution of Sex within Experimental Groups.
The elaborate randomisation procedures described in chapter three failed to produce experimental groups containing similar numbers of male and female subjects. The numbers of male and female subjects within each experimental treatment are shown in Table 4.1.
Table 4.1.

Constitution of Experimental Groups

<table>
<thead>
<tr>
<th>Individual Social Code</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Treatment</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Negative None</td>
<td>-/0</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Positive None</td>
<td>+/0</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Matching None</td>
<td>=/0</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Actual None</td>
<td>*/0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Negative Positive -/+</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Positive Positive +/+</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Matching Positive =/+</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Actual Positive */+</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Negative Negative -/-</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Positive Negative +/-</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Matching Negative =/-</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Actual Negative */-</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>None None Control</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

Total N 22 108 130
The totals of columns in table 4.1 reflects the proportion of male to female students within the college population. It should be noted that the experimental groups which were exposed to positive social treatment had a lower proportion of male subjects than was the case within the other two social treatments.

2. Performance and Expected Score

Figure 4.1 presents graphs of performance and expected score for each experimental group over all four trials. It should be noted that the experimental treatments were applied at the conclusion of trials 2 and 3. Trial 1 was an adaption trial where no feedback was applied and Trial 4 was the final performance measure. The effects of the first application of experimental treatment should therefore be represented in the Trial 3 scores and the effect of the second treatment in the Trial 4 scores.

The Trial 2 scores, which followed the adaption trial, are relatively uncontaminated with the variation which could be expected in view of the novelty of the procedures. Trial 2 scores are therefore regarded as the pre-test scores for all experimental groups.
For all experimental groups over all trials, mean performance scores (solid line) and mean expected scores (dotted line).
Table 4.2.

Mean Performance Scores (P) and Mean S.A.I. (difference between performance and expected mark) for all experimental groups on all trials.

<table>
<thead>
<tr>
<th>Individual Treatment</th>
<th>Social Treatment</th>
<th>P/S</th>
<th>TRIAL I</th>
<th>TRIAL II</th>
<th>TRIAL III</th>
<th>TRIAL IV</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>None</td>
<td>P</td>
<td>2.9</td>
<td>3.1</td>
<td>3.2</td>
<td>3.7</td>
<td>-/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>1.3</td>
<td>1.1</td>
<td>1.3</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>P</td>
<td>2.8</td>
<td>3.8</td>
<td>2.8</td>
<td>3.1</td>
<td>-/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>1.4</td>
<td>1.3</td>
<td>1.2</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>P</td>
<td>3.0</td>
<td>2.7</td>
<td>3.1</td>
<td>4.1</td>
<td>-/-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>0.5</td>
<td>1.4</td>
<td>0.5</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>POSITIVE</td>
<td>None</td>
<td>P</td>
<td>3.2</td>
<td>3.2</td>
<td>3.6</td>
<td>3.3</td>
<td>+/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>0.4</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>P</td>
<td>3.4</td>
<td>3.8</td>
<td>3.9</td>
<td>3.6</td>
<td>+/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>-0.2</td>
<td>0.5</td>
<td>0.7</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>P</td>
<td>3.4</td>
<td>3.9</td>
<td>3.5</td>
<td>4.2</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>0.3</td>
<td>0.7</td>
<td>1.0</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>MATCHING</td>
<td>None</td>
<td>P</td>
<td>2.6</td>
<td>3.3</td>
<td>4.6</td>
<td>3.4</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>1.3</td>
<td>1.2</td>
<td>0.6</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>P</td>
<td>3.0</td>
<td>3.0</td>
<td>3.2</td>
<td>3.3</td>
<td>=/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>0.9</td>
<td>0.9</td>
<td>0.6</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>P</td>
<td>3.4</td>
<td>3.8</td>
<td>3.1</td>
<td>3.6</td>
<td>=/-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>0.8</td>
<td>0.6</td>
<td>1.5</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>ACTUAL</td>
<td>None</td>
<td>P</td>
<td>2.9</td>
<td>2.7</td>
<td>3.4</td>
<td>3.3</td>
<td>*/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>0.7</td>
<td>1.2</td>
<td>0.2</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>P</td>
<td>3.4</td>
<td>3.2</td>
<td>3.8</td>
<td>3.7</td>
<td>*/+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>1.7</td>
<td>1.5</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>P</td>
<td>3.3</td>
<td>3.6</td>
<td>3.4</td>
<td>4.1</td>
<td>*/-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAI</td>
<td>1.3</td>
<td>0.8</td>
<td>0.9</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>NONE</td>
<td>None</td>
<td>P</td>
<td>4.1</td>
<td>3.2</td>
<td>4.2</td>
<td>3.4</td>
<td>CONTROL</td>
</tr>
</tbody>
</table>
3. Performance and Accuracy of Self-assessment

Table 4.2 presents the mean performance scores of each experimental group (as in Fig. 4.1) together with the mean discrepancy between performance score and expected score (S.A.I.) within that group over all four trials. (S.A.I. = expected score minus performance score).
4. **Performance following combinations of Social and Individual Experimental Treatment**

Table 4.3 presents the mean gain scores in performance following experimental treatments. Gain score 1 is the Trial 3 score for one subject, minus the Trial 2 score for that subject and represents performance increment following one application of an experimental treatment combination. Gain 2 score represents the performance increment following the second application of the same treatment combination. (Trial 4 score minus Trial 3 score). Gain score 3 is the total performance increment from 'pretest' to 'post test' (Trial 4 score minus Trial 2 score) following two applications of the same experimental treatment combination.
Table 4.3
Total gain in performance scores following experimental treatments within all experimental groups.

<table>
<thead>
<tr>
<th>Individual Social Treatment</th>
<th>Gain 1</th>
<th>Gain 2</th>
<th>Gain 3</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>-/0</td>
</tr>
<tr>
<td>NEGATIVE Positive</td>
<td>-10</td>
<td>3</td>
<td>-7</td>
<td>-/+</td>
</tr>
<tr>
<td>Negative</td>
<td>4</td>
<td>10</td>
<td>14</td>
<td>-/-</td>
</tr>
<tr>
<td>None</td>
<td>4</td>
<td>-3</td>
<td>1</td>
<td>+/0</td>
</tr>
<tr>
<td>POSITIVE Positive</td>
<td>1</td>
<td>-3</td>
<td>-2</td>
<td>+/-</td>
</tr>
<tr>
<td>Negative</td>
<td>-4</td>
<td>7</td>
<td>3</td>
<td>+/-</td>
</tr>
<tr>
<td>None</td>
<td>13</td>
<td>-12</td>
<td>1</td>
<td>=/0</td>
</tr>
<tr>
<td>MATCHING Positive</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>=/+</td>
</tr>
<tr>
<td>Negative</td>
<td>-7</td>
<td>5</td>
<td>-2</td>
<td>=/-</td>
</tr>
<tr>
<td>None</td>
<td>7</td>
<td>-1</td>
<td>6</td>
<td>*/0</td>
</tr>
<tr>
<td>ACTUAL Positive</td>
<td>6</td>
<td>-1</td>
<td>5</td>
<td>*/+</td>
</tr>
<tr>
<td>Negative</td>
<td>-2</td>
<td>7</td>
<td>5</td>
<td>*/-</td>
</tr>
<tr>
<td>NONE</td>
<td>None</td>
<td>10</td>
<td>-8</td>
<td>2 Control</td>
</tr>
</tbody>
</table>

Gain 1 = Trial 3 score - Trial 2 score (First application of treatment)

Gain 2 = Trial 4 score - Trial 3 score (Second application of treatment)

Gain 3 = Trial 4 score - Trial 2 score (Overall gain following two applications of treatment)
5. The Effects of the Social Treatments upon performance

The three social treatments:

a. 'Positive' (subject informed that performance was above average

b. 'None' (no social information) and

c. 'Negative' (subject informed that performance was below average)

were each applied to four groups of ten subjects in combination with four individual treatments. The three 'social' groupings were therefore equivalent with respect to other experimental conditions.

The total gain scores for these groups (Gain 1 following first application of social treatment, etc.) are presented in table 4.4.

Table 4.4. Social Treatment
Total gain in performance scores following experimental treatments within social treatments.

<table>
<thead>
<tr>
<th>Social Treatment</th>
<th>Gain 1</th>
<th>Gain 2</th>
<th>Gain 3</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>25</td>
<td>-11</td>
<td>14</td>
<td>40</td>
</tr>
<tr>
<td>Positive</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>40</td>
</tr>
<tr>
<td>Negative</td>
<td>-9</td>
<td>29</td>
<td>20</td>
<td>40</td>
</tr>
</tbody>
</table>

Gain 1, 2, 3 as in table 4.3.
The mean performance scores at each trial for the social treatment groups are displayed in Fig. 4.2.

Fig. 4.2.
Social Treatments: mean scores by trials (N=40).

6. The Effects of the Individual Treatment upon performance

The four individual treatments:

a. 'Positive' (subject informed that performance score was one point more than expected score) "+"

b. 'Negative' (one point less than expected) "-"

c. 'Matching' (equal to expectation) "="

d. 'Actual' (actual performance score) "*"

-44-
were each applied to three groups of ten subjects, equivalent with respect to other experimental conditions as in the social treatments (5 above). The total gain scores for these groups are presented in table 4.5.

Table 4.5. Individual Treatments
Total gain in performance scores following experimental treatments within individual treatments.

<table>
<thead>
<tr>
<th>Individual treatment</th>
<th>Gain 1</th>
<th>Gain 2</th>
<th>Gain 3</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>-5</td>
<td>18</td>
<td>13</td>
<td>30</td>
</tr>
<tr>
<td>Positive</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Matching</td>
<td>8</td>
<td>-6</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Actual</td>
<td>11</td>
<td>5</td>
<td>16</td>
<td>30</td>
</tr>
</tbody>
</table>

Gain 1, 2, 3 as in table 4.3.

The mean performance scores at each trial for the individual treatments are displayed in Fig. 4.3.

Table 4.6. Gain scores; All Treatments
Total gain in performance scores following experimental treatments; all treatment groups combined.

<table>
<thead>
<tr>
<th>Gain 1</th>
<th>Gain 2</th>
<th>Gain 3</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>All groups</td>
<td>15</td>
<td>18</td>
<td>33</td>
</tr>
</tbody>
</table>

Gain 1, 2, 3 as in Table 4.3.

-45-
7. Interaction effects upon performance; Social and Individual treatments.

The interaction effect between social and individual treatments are displayed in Figs. 4.4 and 4.5. Initial differences in pretest performance (Trial 2) are 'corrected' by the use of gain scores.
Fig. 4.4.

Interaction effects I. Individual Treatment Gain scores (Trial 4 - Trial 2) against 'levels' of social treatment.
Fig. 4.5.

Interaction effects II. Social treatment gain scores (Trial 4 - Trial 2) against levels of individual treatment.
SECTION B. Inferential statistics

1. Treatment Combinations and Performance.
The statistical significance of the performance differences between the means of all thirteen experimental groups was tested by Analysis of Covariance (Garrett 1958, P.296), adjusting differences in final performance scores (Trial 4) for initial differences in pretest scores (Trial 2). Table 4.7 presents a summary of this analysis.

Table 4.7.
Summary of Analysis of Covariance. All treatment groups (after Garrett P.296).

I. Analysis of Variance of Trial 2 and Trial 4 scores

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SSx (Trial 2)</th>
<th>SSy (Trial 4)</th>
<th>Msx</th>
<th>Msy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between means</td>
<td>12</td>
<td>10.7</td>
<td>14.8</td>
<td>1.72</td>
<td>1.23</td>
</tr>
<tr>
<td>Within groups</td>
<td>117</td>
<td>110.3</td>
<td>238.2</td>
<td>1.88</td>
<td>2.04</td>
</tr>
</tbody>
</table>

Fx (Trial 2) = \frac{1.72}{1.88} = 0.9148 \text{ N.S. at df } 12/117

Fy (Trial 4) = \frac{1.23}{2.04} = 0.6029 \text{ N.S.}

II. Analysis of Covariance

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SSx</th>
<th>SSy</th>
<th>SSxy</th>
<th>SSyx</th>
<th>MSyx</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between means</td>
<td>12</td>
<td>20.7</td>
<td>14.8</td>
<td>1.5</td>
<td>14.69</td>
<td>1.224</td>
</tr>
<tr>
<td>Within groups</td>
<td>116</td>
<td>220.3</td>
<td>238.2</td>
<td>53.5</td>
<td>225.21</td>
<td>1.941</td>
</tr>
</tbody>
</table>

Fy.x = \frac{1.244}{1.941} = 0.641 \text{ N.S.}
The Fx on Trial 2 means indicates that the randomisation and adaptation procedures were reasonably successful. No significant differences were found amongst the performance of the groups when combinations of social and individual treatments were applied.

2. Social Treatment, Individual Treatment, Interaction and Performance.

The possible confounding of the effects of individual treatments with social treatments was examined by two-way analysis of variance using gain scores from Trials 2 to 4. The gain scores were 'coded' by adding a constant of 10 (Guilford & Fruchter 1973, P.237). A summary of this analysis is presented in table 4.8.
Table 4.8.
Summary of two-way Analysis of Variance (four levels of individual treatment and three levels of social treatment).

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>M.S.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between means</td>
<td>33.9</td>
<td>11</td>
<td>3.082</td>
<td>N.S.</td>
</tr>
<tr>
<td>Within means</td>
<td>339.1</td>
<td>108</td>
<td>3.114</td>
<td></td>
</tr>
<tr>
<td></td>
<td>373</td>
<td>119</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Further Analysis:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>M.S.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Social</td>
<td>7.1</td>
<td>2</td>
<td>3.05</td>
<td>N.S.</td>
</tr>
<tr>
<td>Between Individual</td>
<td>3.7</td>
<td>3</td>
<td>1.23</td>
<td>N.S.</td>
</tr>
<tr>
<td>Interaction</td>
<td>23.0</td>
<td>6</td>
<td>3.83</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

No significant differences were found in this analysis.

A further two-way analysis, excluding the 'actual' individual treatment groups, was carried out and again no significant differences were found between the remaining groups.

Further analyses could have been carried out to test the differences in means on Gain Scores I (one application of treatment), and Gain Score II. These analyses had not been planned prior to the experiment (see assumptions with respect to additive effect of...
incentive applications) and carried the danger of a chance significant result. A compromise was effected by reducing the error within measures of individuals performance over all four trials.

It could be argued that Trial 1 and Trial 2 scores are two observations of the subjects ability to perform a task prior to experimental treatment. These observations are subject to measurement error and the mean of these two observations could be held to be an improved estimate of "true" ability prior to motivation. (This argument of course, ignores the original purpose of Trial 1, that of providing an adaptation period).

Similarly, the mean of the Trials 3 and 4 scores could be held to represent an improved estimate of the Post-Motivation "true" ability. An analysis of variance of the gain score derived from "Pre-Motivation" and "Post-Motivation" mean scores was carried out. No significant differences were found. A summary of this analysis is presented in Table 4.9.
Table 4.9.
Summary of two-way Analysis of Variance; (Four levels of individual treatment x three levels of social treatment; Gain scores: Post-Motivation Mean Score - Pre-Motivation Mean Score.)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between means</td>
<td>67</td>
<td>11</td>
<td>6.09</td>
<td>N.S.</td>
</tr>
<tr>
<td>Within means</td>
<td>748</td>
<td>108</td>
<td>6.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>815</td>
<td>119</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Further Analysis:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Social</td>
<td>12.7</td>
<td>2</td>
<td>6.35</td>
<td>N.S.</td>
</tr>
<tr>
<td>Between Individual</td>
<td>40.0</td>
<td>3</td>
<td>15.30</td>
<td>N.S.</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.3</td>
<td>6</td>
<td>0.30</td>
<td>N.S.</td>
</tr>
</tbody>
</table>
SECTION C. Experimental Treatment and Self-assessment of Performance

As noted earlier, the major purpose of laboratory study had been the investigation of the effects of experimental treatments upon performance. The experimental conditions also provided an opportunity to observe the relationship between performance score and expected score defined in chapter 2 as the Self-Assessment Index (S.A.I.). The S.A.I. can also be regarded as an additional dependent variable within the total experimental situation and was treated as such for the following analyses. The sequence of analysis is the same as that followed in Section A, subsections 5 to 7.

1. Effects of the Social Treatments upon S.A.I.

The mean S.A.I. at each trial for the social treatment groups are displayed in Figure 4.6.
Fig. 4.6.
Social Treatment and Self-Assessment Index: mean S.A.I. by trials (three groups of N 40).

Effects of the Individual treatments upon S.A.I.
The mean S.A.I. at each trial for the individual treatment groups are displayed in Fig.4.7.
3. **Interaction Effects upon S.A.I.; Social and Individual Treatments.**

The interaction effects between social and individual treatments are displayed in Figs. 4.8 and 4.9. As in subsection 7 of Section A, 'gain' scores, (S.A.I. on Trial 4 minus S.A.I. on Trial 2) are used.
Fig. 4.8. Interaction effects I: Mean gain in S.A.I. (Trial 4-2). Individual treatments against levels of Social Treatment.
Since the differences in SAI between the groups did not deviate appreciably from differences in performance observed earlier (Sections A and B) the statistical significance of differences in SAI are assumed to be also below an acceptable level, and no further inferential analyses were carried out on the data.
CHAPTER 5

THE LABORATORY STUDY: CONCLUSIONS

The results of the laboratory study show that there are no statistically significant differences in the performance of groups classified according to experimental treatment; and further, that there was no statistically significant interaction effect between social and individual treatment.

The trends in the results indicate, however, (from extrapolation), that significant differences might have occurred, had an additional trial (and additional treatment) been part of the design. Similarly, there are indications that social information can affect performance, particularly when an individual has his/her expectation disconfirmed negatively (see fig. 4.5).

The implications of these findings are;

(a) There are no performance effects to be expected from this type of feedback in classrooms, or
(b) The laboratory design was not an adequate representation of classroom conditions.

If no effects are observed in further classroom experiments this would confirm (a) above, and further, would indicate that the findings of previous work were obtained by the operations of chance alone. However, evidence from within the laboratory study indicates that the design was inadequate in some respect. The fact that no significant differences occurred, even with the inclusion of a control group receiving no feedback, indicates that the design was
inadequate to observe an effect which has been demonstrated repeatedly in laboratory conditions: that performance following knowledge of results is superior to performance following no knowledge of results.

"The facilitative effect of knowledge of results is one of the best established findings in the research literature..." (Locke E.A. et al, 1968).

A possible defect of design can be illustrated by a further quote from the paper;

".... knowledge of results.... may also affect motivation, depending on the degree to which S values being correct and wants to improve his performance."

With respect to the observations of Self-Assessment Index, it is interesting to note that there appears to be a negative correlation between SAI and performance, reflected in the comparison of some trends displayed in;

Fig. 4.2 and Fig. 4.6  
Fig. 4.3 and Fig. 4.7

Further examination of this relationship appeared to be a worthwhile feature of an extended classroom study.
PART II

THE MAIN STUDY
CHAPTER 6
THE BACKGROUND TO THE MAIN STUDY

INTRODUCTION

As noted in Chapter 2 of the present study, previous experimental investigation of effects of discrepancies between marks expected and marks received (Pickup 1967, Pickup and Anthony 1968) had produced findings which indicated two different types of further investigation, a laboratory investigation, and extended classroom investigation. Part II of the present study describes this further research, conducted in classroom contexts and designed to elucidate and extend the findings of previous work.

THE BACKGROUND TO THE MAIN STUDY

The previous work had demonstrated that there was no significant difference in classroom performance between groups exposed, on a single occasion to one of the three experimental treatments of:

(a) Receiving a mark greater than expected
('positive' treatment)
(b) Receiving a mark less than expected
('negative' treatment) and,
(c) Receiving a mark equal to that expected
('matching' treatment).

Further findings indicated that these treatments appeared to have some effects contingent upon the ability level of the individual, and particularly, that these effects tended to reverse as treatments were applied at differing ability levels. An attempt to explain non-significant differences and the reversal effect in terms of a new intervening variable was
made in the laboratory investigation described in Part I. As reported there, no significant differences in performance were found between groups exposed to repeated experimental treatment where 'social' information was controlled as an experimental variable. While the trends were in expected directions, the effects were not so large as to require the inclusion of a 'social' variable in further classroom experiments. The laboratory investigation had also demonstrated that a reduction of experimental error was insufficient to permit significant performance effects to emerge. The reasons that the previous work had failed to demonstrate a single incentive effect for a particular experimental treatment could, on the basis of the laboratory experience, be (i) concerned with the frequency of application of experimental treatment (trends were proceeding towards significant differences after successive applications of experimental treatment) or (ii) an uncontrolled variable (other than social information) intervening between independent variable and dependent variable. The best candidate as the uncontrolled variable of previous work was now seen to be the subjects past history of experience in similar feedback situations.

To a certain extent both (i) and (ii) above are related. If the effects upon performance of successive applications of experimental treatment are studied, the effects in later trials are partly dependent upon the experience of the subject in earlier trials. By allocating a subject to an experimental group whose history of experience is controlled and standardised it is supposed that 'random' differences in the
subject's previous history become less potent over successive trials, allowing a possible single effect to appear. An experiment conducted along these lines would also provide, at its conclusion, a set of groups whose history had been controlled for some time, and who could then be studied for the effect of introducing a change of experimental treatment.

A limitation of earlier work had been identified (Pickup 1967, P23) as similar to a limitation of the designs which had been employed in the allied field of Praise and Blame. Kennedy and Willcutt, (1964), reviewing work in the Praise and Blame area, state that further classification would seem to require a functional design which

"..... in some way controls the reinforcement history of the subjects".

In Pickup and Anthony (1968) it was shown that low-scoring pupils were especially liable to over-estimate their performance relative to the teachers mark. If this over-estimation was a feature of the subjects' history prior to the introduction of experimental treatment it would follow that the application of positive treatment would be the opposite of feedback which tends to occur in normal classroom practice. To a lesser extent this argument could be extended to include high-scoring pupils; who generally tended to under-estimate their mark and whose performance tended to improve following negative treatment.

As in the laboratory study, any further experiment would have to be designed so as to include features concerned with (i) elucidating earlier findings, (ii) with extending the scope of the research so as to include comparison of
experimental treatment with 'normal' classroom feedback, and
(iii) further observation of the expected score of the
individual, relative to teacher's mark.

THE MAIN STUDY: RATIONALE AND DISCUSSION.

This study had the same general aims of previous work and
the laboratory study: the investigation of the phenomena
existing in a classroom, when, following the performance of a
task, a mark is returned to a pupil by a teacher. These
general aims and the primary assumptions are described in more
detail in Chapter 2. The main study required, however, a
fundamentally different mode of investigation from that
employed in the laboratory study. The aim of generalisability
to normal classroom contexts (which was a feature of previous
work) was seen as a necessary condition of an investigation
conducted 'in the field'. This aim imposed a number of
constraints upon experimental design, not least in the
measurement of performance. The comparatively poor reliability
of performance measures of previous work indicated that
classroom tests would not be a good instrument for observing
small effects. However, in assessing the worth of an
incentive technique it could be argued that even relatively
poor instrumentation should be able to detect a consistent
effect, if the effect were of practical significance. It was
decided that, in any further experiment, data would be
accepted from all classroom tests, standardising scores as in
previous work. If no significant effects appeared, a
retrospective sampling of data only from tests with a
reliability index better than 0.6 could then be studied for
the existence of statistically significant effects.
The size of effect noted in previous work, together with the high experimental mortality on only two test occasions (caused by absence from school) indicated that a much larger sample than that studied previously would be required in any extended experiment.

A further constraint of extended experiment was the credibility of repeated experimental discrepancies of a fixed size. Previous work had not demonstrated significant effects resulting from experimental discrepancies of 3 marks (15% of the possible range of scores). It was considered that this discrepancy was too large to be accepted as normal by the subjects, when repeated over a series of trials. The number of trials and the size of experimental discrepancy were factors which required some balancing. It was decided that an experimental discrepancy of 10% (between marks received and expected) over three trials was the most that could be expected to be credible to the subjects; who should of course be unaware that an experiment was taking place.

The final design was influenced by the foregoing factors so that the main study took the form of a pre-test, followed by three applications of experimental treatment, followed again by three applications of a different treatment. Four experimental treatments were employed. These may be classified as:

A. Fixed discrepancy (between mark returned and mark expected); 1. 'Positive' (10% more than expected). 2. 'Negative' (10% less than expected).

B. No discrepancy; 3. 'Matching' (same as expected).

C. Variable discrepancy; 4. 'Actual' (return of actual Performance score).
The main study was designed as an extension to previous work in the following respects:

1. Increase in sample size.
2. Inclusion of a 'normal' (Actual) feedback treatment as an experimental variable.
3. Increased frequency of experimental treatment.
4. Observation of effects subsequent to controlled history of feedback.

This general design also provided opportunity to:

5. Replicate certain conditions of previous work so as to confirm or disconfirm the reversal effect noted there.
6. Observe further the relationship between mark expected and mark received, specifically the relationship between self-assessment and exterior assessment, by means of S.A.I.
7. Observe the effect of experimental treatment upon S.A.I.

THE RESEARCH QUESTIONS

1. Are there any significant differences in the incentive action of four differing types of feedback over repeated applications of such feedback?
2. What are the effects of introducing an experimental treatment which differs from that experienced by a subject on three previous occasions?
3. What are the effects of experimental treatments upon Self-Assessment Index?
4. When data from the first two trials are examined, are the findings of previous work (concerning differential effects of treatment at ability levels) confirmed or disconfirmed in a large sample?

5. Within all the above, are there any significant differences between the sexes?
CHAPTER 7

THE MAIN STUDY : METHOD

Introduction

The purpose of the main study was to examine the effects of four types of feedback upon classroom performance and self-assessment in secondary schools. Each participating teacher was asked to devise, administer and mark seven tests of normal classwork and to apply experimental procedures according to a detailed specification supplied by the experimenter (see Appendix F). Each participating class was divided into four experimental groups at random within ability levels and each group was exposed to a sequence containing two types of feedback over the seven tests. Measures of performance and of expected score (self-assessment) were made throughout.

The four types of feedback were represented as the following experimental treatments:

1. "Actual" (code: "*"). Here, pupils were returned the performance mark determined by the teachers' marking criteria and not distorted for motivational purposes. The discrepancy between this mark and the pupils expected mark could therefore vary within the limits of test range.

2. "Matching" (code "="). Here, pupils were returned a mark which exactly matched the pupils expected score (regardless of actual performance). There was thus no discrepancy between mark returned and the mark expected.
3. "Negative" (code "••"). Pupils were here returned a mark which was 2 marks (10% of range of possible score) less than they had expected (again, regardless of actual performance score).

4. "Positive" (code "+"). Pupils were here returned a mark which was 2 marks more than they had expected.

Each experimental group was exposed to three applications of a particular experimental treatment followed by three applications of the 'opposite' treatment. The experimental groups are thus defined as:

1. Actual followed by Matching; code; "*(=)".
2. Matching followed by Actual; code; "=(*)".
3. Positive followed by Negative; code; "+(−)".
4. Negative followed by Positive; code; "−(+)".

Standardisation of test scores

The raw performance scores from each test were standardised by conversion into "T" scores (Garrett 1958 pp 314-318, Table G). The percentile rank of each score from a single test was computed and converted into a "T" score; so that each test provided data with a mean of 50 and an SD of 10. "T" scores from different tests are said to be comparable when the distribution of the trait measured is normal in the population.

The performance measures

Participating teachers discussed the content and type of test to be administered to classroom groups with the experimenter (see Experimental Briefing below) and with minor adjustments (taking age, ability and curriculum into
account) agreed to follow the broad specifications in the "Information for participating teachers" pamphlet (see Appendix F, construction of the tests). These specifications are repeated here:

1. Tests should be constructed of easily scored objective-type items which require no subjective interpretation.

2. Each test should be as independent of each other as possible in terms of the material to be tested.

3. Each test should be constructed so as to give every pupil an opportunity of gaining some marks, but extreme difficulty in gaining full marks.

4. The pupils may be given the impression that they may gain or lose marks in some subjective way.

The purpose of No. 4 above was to 'camouflage' the experimental distortion of marks. Most participating teachers used the simple admonition "Neatness counts!". Pupils' answers were to be on loose paper and not returned to pupils until conclusion of the experiment.

The following are the approximate proportional representations of the various school subjects included in the main study.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>44%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>22%</td>
</tr>
<tr>
<td>Modern Languages</td>
<td>18%</td>
</tr>
<tr>
<td>Social Studies</td>
<td>11%</td>
</tr>
<tr>
<td>General Science</td>
<td>4%</td>
</tr>
</tbody>
</table>
The measure of expected score

At the conclusion of each test in the experimental series the participating teacher distributed additional slips of paper. Each pupil was asked to record his/her name, class and the date on this slip and to record the mark out of 20 which they believed their work to merit. Standardised instructions (see Appendix F) regarding this measure were applied by the participating teacher.

This slip of paper (termed the pupils' assessment slip) was used to return a mark to the pupil from the teacher, thus serving as a vehicle for transmission of experimental treatment.

The experimental briefing

The twenty teachers who had expressed interest in participation were invited to meetings with the experimenter. At all these meetings the teachers were advised that they could and should cease to apply experimental treatment to any child who appeared to be in danger of distress. (Two such exclusions were finally reported to the experimenter). This was achieved by allocating such children to the "No Treatment" category (described later in this chapter). The full experimental procedures were discussed, together with the selection of test content and the selection of suitable classes/groups in which to apply the experiment. Four teachers withdrew at this point and sixteen teachers elected to participate.
These participating teachers were informed of the background to the study, and that the previous work had not demonstrated a single incentive effect for any experimental treatment. They were then supplied with sufficient special record forms for the number of children involved. Detailed discussion of procedures ensued, and finally, teachers were supplied with stamped addressed envelopes for the return of the record forms to the experimenter. Participating teachers were encouraged to contact the experimenter if any difficulties emerged during the period of the study.

Allocation to experimental treatment

Prior to the experimental briefings the experimenter had prepared sufficient experimental record forms (see example in Appendix F) so that random allocation to experimental treatment within ability levels (as measured by the first test of the series) could be achieved. The record forms were printed with two blank columns (headed Treatment Code I and II) by which a particular experimental treatment could be applied to an individual occupying a row position. By means of a specially constructed 4-sided die the experimenter allocated an experimental treatment to the top row of the record form. He then allocated at random one of the three remaining treatments to the second row, and one of the two remaining treatments to the third row. The fourth row was allocated the remaining treatment. The next 'level' of four rows was allocated from bottom to top. These procedures were commenced on a form labelled Sheet A, and continued onto Sheets B and C.
At the conclusion of the first test in the series the participating teacher was asked to mark the test scripts and to place them in rank order of score from top to bottom. Tied ranks were to be ordered at random. The participating teacher was asked to enter the names of individuals onto the record form, thus allocating a 'row' treatment to an individual.

The record form

As stated above, the record form was used as a method of randomly allocating treatment to individuals within ability levels. An example form is in Appendix F. The forms were used to record (i) the marks gained by pupils on each test, (ii) the expected score for that test, and (iii) the mark which was subsequently to be returned to the pupils. They also presented the treatment codes for each row, thus providing the participating teacher with the information required for rapid determination of the mark to be returned. (Note, the "Actual" treatment group was described to teachers as the "Control" group, and the code used to identify this treatment was a tick.

The special procedures

These were procedures to be adopted (i) in the case of absence from a particular test or tests or (ii) in allocation to "No Treatment". The no treatment category is a feature of the laboratory study and previous work and occurs when treatment can not be applied because of extreme expected scores. The procedures are fully described in Appendix F,
and could be applied to other children as described under "experimental briefings" above.

The treatment of data

When all the participating teachers had returned the record forms to the experimenter the data was scrutinised to confirm that a correct 'returned' mark had been allocated for each experimental condition. The data from eleven children was rejected because of error of this type. The data was then transferred to punched cards for computer analysis and the punched cards were 'verified' to eliminate punching error. Each computer analysis contained procedures designed to eliminate individuals whose scores fell beyond legitimate ranges, and thus also served to eliminate individuals with coded values representing "Absence" or "No Treatment".

The initial sample

The head teachers of twenty schools were asked if they would be prepared to;

(a) Permit research of this type to be conducted within their school, and if so;
(b) To inform their staff of the broad purpose of the research and to ask if any member of the staff would be interested in participating in such research, and if any member of staff expressed interest to;
(c) Arrange a meeting between interested staff and the experimenter.

Nine head teachers agreed, and after the meetings (and experimental briefings), sixteen teachers in eight schools
agreed to carry out the experimental procedures with 27 existing school groups or classes.

The eight schools were;

1. A girls Grammar school in London Borough
2. Two Comprehensive schools in Central London
3. A Junior High school in a second London Borough
4. A Middle school in a third London Borough
5. A non-selective Secondary school in a fourth London Borough
6. A girls Grammar school in Central London
7. A non-selective Secondary school in Central London

Classes within these schools were selected to represent "average ability" pupils within the school population (as far as the participating teachers' timetable permitted). The sample of pupils commencing the experimental procedures (and for whom legitimate data was available) was as follows:

<table>
<thead>
<tr>
<th>Boys</th>
<th>351</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td>418</td>
</tr>
<tr>
<td>Total</td>
<td>769</td>
</tr>
</tbody>
</table>

Post-Hoc Sampling

The data which would provide the most accurate conclusions about the effects of experimental treatment could be drawn only from pupils who had been present at every test session. While some experimental 'mortality' had been anticipated, the actual mortality was very high. Illustrative figures are as follows:
Table 7.1 Experimental mortality: Performance

<table>
<thead>
<tr>
<th></th>
<th>Number Present for tests 1 and 2</th>
<th>Number Present for all tests</th>
<th>Percentage 'Loss of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>boys</td>
<td>319</td>
<td>213</td>
<td>33.23</td>
</tr>
<tr>
<td>girls</td>
<td>394</td>
<td>258</td>
<td>34.52</td>
</tr>
<tr>
<td>TOTAL</td>
<td>713</td>
<td>471</td>
<td>33.94</td>
</tr>
</tbody>
</table>

Table 7.1 shows that over a third of those pupils commencing the experiment were absent (or otherwise excluded, less than 4%) for some test later in the series.

A retrospectively drawn sample, consisting only of pupils who had been present for all the seven tests would be plainly unrepresentative of the normal school population. However, the inclusion of pupils who had not been present at all test sessions (and had thus been exposed to reduced frequency of treatment) would present difficulties in interpreting frequency and order effects of treatment.

It was decided to sample the data in two ways, each thought appropriate to the dependent variable under consideration;

(a) Investigation of Performance

Since the major interests here were, (i) investigation of the effects of the first treatment type in the series (particularly involving the first two trials) and (ii) investigation of treatment effect following controlled history of 'opposite' treatment; it was considered that a progressive reduction of sample would be acceptable. This
was achieved in the following stages; (i) excluding all individuals not present for both test I and test II, (ii) excluding individuals absent for a later test only at that point in the series (permitting the use of their performance scores prior to absence) (iii) excluding individuals allocating to 'No Treatment' (4% of initial sample) again, only at the point that this event occurred.

(b) Investigations of SAI

It was decided here to use all the available data for observations concerning self-assessment on any one trial, and to accept some experimental error with respect to an order or frequency effect of treatment.

It was further decided to standardise all test scores by computing the percentile rank of all pupils present for that test, whether or not the resultant T score data from some pupils would be used in analysis. The percentile rank of scores therefore reflect relative performance within a normal attendance group of pupils.

Null hypotheses and Data Analysis

The research questions of chapter 6 were converted (where appropriate) to Null hypotheses as follows:

(i) There will be no significant difference in performance, as measured by T. scores derived from classroom tests, between groups exposed to four differing forms of feedback, or between groups of boys and girls exposed to these forms of feedback.
(ii) There will be no significant difference in Self-Assessment Index on any classroom test of a series, between groups exposed to four forms of feedback, or between boys and girls exposed to these forms of feedback.

(iii) There will be no significant correlation between the mean pre-test score of a matched-pair (Ss receiving positive and negative treatment within the same ability level) and the difference in gain score (Trial II - Trial I) between the subjects in that matched-pair.

(iii) above is a replication of procedures followed in previous work, which led to the finding of the 'reversal effect' described earlier.

The randomisation within blocks (or ability levels) employed in this study could permit the use of procedures which would reduce the error term; eg two-way analysis of variance with ability level as the 'row' variable. However, the experimental mortality which would be induced by such a procedure (elimination of a block of four subjects occasioned by the absence of one subject) was considered unacceptable. Accordingly a simple analysis of variance was selected, as a conservative test of these hypotheses.

The effects of treatment upon ability levels was of course implicit in hypothesis (iii) above, and a product-moment correlation was selected for this analysis.
CHAPTER 8

THE MAIN STUDY: RESULTS.

Introduction

The results of this study fall into three major categories;

A. Experimental Treatment and Performance.
B. Experimental Treatment and Self-Assessment.
C. Self-Assessment and Performance.

Results pertaining to these categories are presented in separate sections below.

SECTION A. Experimental treatment and performance

Mean T. scores for each experimental treatment group, and for boys and girls separately within treatment were computed for each experimental trial. The data from each experimental trial was further subjected to two analyses of variance, the first concerned with performance within experimental treatment groups, the second with boys and girls separately within treatments.

Table 8.1 presents the mean scores at each trial for all groups and summaries of the analyses of variance associated with each trial. (As noted in the previous chapter, the sample size reduces progressively throughout the seven tests). The table shows that the null hypothesis concerning differences between experimental treatment groups can be rejected at Trials III and IV, that is after two and three applications of a single experimental treatment. The non-significant F on trial
I indicates that the randomisation procedures were effective in selecting groups of equal ability.

The significance of differences between groups selected by sex and experimental treatment show that the null hypothesis may be rejected in the first half of the experiment, but differences fail to reach a significant level in the second half of the experiment.

The mean T scores for each treatment group are displayed graphically in Fig. 8.1. The mean T scores for boys and girls within treatment are displayed in Fig. 8.2.
### Table R.1: Experimental treatments and Performance. T-Scores by sex, all trials

<table>
<thead>
<tr>
<th></th>
<th>Trial I</th>
<th></th>
<th>Trial II</th>
<th></th>
<th>Trial III</th>
<th></th>
<th>Trial IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S D</td>
<td>N</td>
<td>Mean</td>
<td>S D</td>
<td>N</td>
<td>Mean</td>
<td>S D</td>
</tr>
<tr>
<td></td>
<td>boys</td>
<td></td>
<td></td>
<td>girls</td>
<td></td>
<td></td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>+(-)</td>
<td>48.0</td>
<td>8.4</td>
<td>70</td>
<td>48.3</td>
<td>9.4</td>
<td>66</td>
<td>48.6</td>
<td>9.4</td>
</tr>
<tr>
<td>all</td>
<td>49.4</td>
<td>8.9</td>
<td>175</td>
<td>49.1</td>
<td>9.1</td>
<td>169</td>
<td>49.5</td>
<td>9.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Trial I</th>
<th></th>
<th>Trial II</th>
<th></th>
<th>Trial III</th>
<th></th>
<th>Trial IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S D</td>
<td>N</td>
<td>Mean</td>
<td>S D</td>
<td>N</td>
<td>Mean</td>
<td>S D</td>
</tr>
<tr>
<td></td>
<td>boys</td>
<td></td>
<td></td>
<td>girls</td>
<td></td>
<td></td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>-(+)*</td>
<td>49.7</td>
<td>8.7</td>
<td>90</td>
<td>47.4</td>
<td>9.2</td>
<td>77</td>
<td>50.3</td>
<td>9.2</td>
</tr>
<tr>
<td>all</td>
<td>50.0</td>
<td>9.1</td>
<td>179</td>
<td>52.4</td>
<td>9.6</td>
<td>100</td>
<td>51.6</td>
<td>10.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Trial I</th>
<th></th>
<th>Trial II</th>
<th></th>
<th>Trial III</th>
<th></th>
<th>Trial IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S D</td>
<td>N</td>
<td>Mean</td>
<td>S D</td>
<td>N</td>
<td>Mean</td>
<td>S D</td>
</tr>
<tr>
<td></td>
<td>boys</td>
<td></td>
<td></td>
<td>girls</td>
<td></td>
<td></td>
<td>all</td>
<td></td>
</tr>
<tr>
<td>(*)</td>
<td>49.6</td>
<td>10.4</td>
<td>80</td>
<td>50.0</td>
<td>10.6</td>
<td>77</td>
<td>47.9</td>
<td>9.7</td>
</tr>
<tr>
<td>all</td>
<td>50.0</td>
<td>9.6</td>
<td>179</td>
<td>49.7</td>
<td>9.1</td>
<td>97</td>
<td>49.3</td>
<td>8.3</td>
</tr>
</tbody>
</table>

For all:

<table>
<thead>
<tr>
<th></th>
<th>Trial I</th>
<th></th>
<th>Trial II</th>
<th></th>
<th>Trial III</th>
<th></th>
<th>Trial IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SS between</td>
<td>62.58</td>
<td>3</td>
<td>136.13</td>
<td>3</td>
<td>775.71</td>
<td>3</td>
<td>700.46</td>
<td>3</td>
</tr>
<tr>
<td>df</td>
<td>3</td>
<td></td>
<td>45.38</td>
<td></td>
<td>258.57</td>
<td></td>
<td>236.49</td>
<td></td>
</tr>
<tr>
<td>MS (bet)</td>
<td>20.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS within</td>
<td>64416.8</td>
<td>709</td>
<td>61889.87</td>
<td>691</td>
<td>57221.86</td>
<td>641</td>
<td>54032.65</td>
<td>596</td>
</tr>
<tr>
<td>df</td>
<td>709</td>
<td></td>
<td>89.57</td>
<td></td>
<td>89.27</td>
<td></td>
<td>90.66</td>
<td></td>
</tr>
<tr>
<td>MS (with)</td>
<td>90.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.23</td>
<td></td>
<td>0.5</td>
<td></td>
<td>2.90</td>
<td></td>
<td>2.61</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>NS</td>
<td></td>
<td>NS</td>
<td></td>
<td>.05</td>
<td></td>
<td>.05</td>
<td></td>
</tr>
</tbody>
</table>

For boys and girls groups:

<table>
<thead>
<tr>
<th></th>
<th>Trial I</th>
<th></th>
<th>Trial II</th>
<th></th>
<th>Trial III</th>
<th></th>
<th>Trial IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SS bet.</td>
<td>1270.05</td>
<td>7</td>
<td>1411.44</td>
<td>7</td>
<td>1775.02</td>
<td>7</td>
<td>1240.23</td>
<td>7</td>
</tr>
<tr>
<td>df</td>
<td>7</td>
<td></td>
<td>201.63</td>
<td></td>
<td>253.57</td>
<td></td>
<td>177.18</td>
<td></td>
</tr>
<tr>
<td>MS (bet)</td>
<td>181.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS within</td>
<td>63209.34</td>
<td>705</td>
<td>60614.56</td>
<td>687</td>
<td>56222.55</td>
<td>637</td>
<td>53501.87</td>
<td>792</td>
</tr>
<tr>
<td>df</td>
<td>705</td>
<td></td>
<td>88.23</td>
<td></td>
<td>88.26</td>
<td></td>
<td>90.37</td>
<td></td>
</tr>
<tr>
<td>MS (with)</td>
<td>80.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2.02</td>
<td></td>
<td>2.29</td>
<td></td>
<td>2.87</td>
<td></td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>&lt;.05</td>
<td></td>
<td>&lt;.05</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>
### TRIAL IV

<table>
<thead>
<tr>
<th>T</th>
<th>N</th>
<th>S D</th>
<th>Mean</th>
<th>S D</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10.1</td>
<td>59</td>
<td>49.3</td>
<td>10.2</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>9.8</td>
<td>85</td>
<td>51.0</td>
<td>9.8</td>
<td>75</td>
</tr>
<tr>
<td>1</td>
<td>9.9</td>
<td>144</td>
<td>50.3</td>
<td>10.0</td>
<td>127</td>
</tr>
<tr>
<td>5</td>
<td>9.9</td>
<td>79</td>
<td>48.5</td>
<td>10.2</td>
<td>69</td>
</tr>
<tr>
<td>5</td>
<td>9.3</td>
<td>69</td>
<td>51.0</td>
<td>7.8</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>9.7</td>
<td>148</td>
<td>49.7</td>
<td>9.2</td>
<td>132</td>
</tr>
<tr>
<td>5</td>
<td>8.0</td>
<td>69</td>
<td>49.1</td>
<td>9.1</td>
<td>63</td>
</tr>
<tr>
<td>0</td>
<td>9.3</td>
<td>83</td>
<td>51.5</td>
<td>8.2</td>
<td>77</td>
</tr>
<tr>
<td>9</td>
<td>8.8</td>
<td>152</td>
<td>50.5</td>
<td>8.7</td>
<td>140</td>
</tr>
<tr>
<td>1</td>
<td>10.2</td>
<td>69</td>
<td>51.3</td>
<td>10.1</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>9.2</td>
<td>87</td>
<td>50.1</td>
<td>9.1</td>
<td>84</td>
</tr>
<tr>
<td>0</td>
<td>9.6</td>
<td>156</td>
<td>50.7</td>
<td>9.6</td>
<td>151</td>
</tr>
</tbody>
</table>

### TRIAL V

<table>
<thead>
<tr>
<th>T</th>
<th>N</th>
<th>S D</th>
<th>Mean</th>
<th>S D</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.7</td>
<td>12.1</td>
<td>44</td>
<td>49.1</td>
<td>9.7</td>
<td>63</td>
</tr>
<tr>
<td>51.6</td>
<td>8.3</td>
<td>65</td>
<td>51.0</td>
<td>7.8</td>
<td>56</td>
</tr>
<tr>
<td>50.0</td>
<td>10.2</td>
<td>109</td>
<td>50.0</td>
<td>9.0</td>
<td>119</td>
</tr>
<tr>
<td>45.9</td>
<td>11.7</td>
<td>57</td>
<td>50.8</td>
<td>11.0</td>
<td>53</td>
</tr>
<tr>
<td>48.3</td>
<td>11.6</td>
<td>110</td>
<td>50.6</td>
<td>11.5</td>
<td>120</td>
</tr>
<tr>
<td>49.6</td>
<td>11.9</td>
<td>55</td>
<td>51.3</td>
<td>11.1</td>
<td>65</td>
</tr>
<tr>
<td>50.6</td>
<td>11.5</td>
<td>120</td>
<td>50.9</td>
<td>9.8</td>
<td>57</td>
</tr>
<tr>
<td>49.8</td>
<td>9.4</td>
<td>75</td>
<td>50.3</td>
<td>9.6</td>
<td>132</td>
</tr>
</tbody>
</table>

### TRIAL VI

<table>
<thead>
<tr>
<th>T</th>
<th>N</th>
<th>S D</th>
<th>Mean</th>
<th>S D</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>638.21</td>
<td>7</td>
<td>1548.93</td>
<td>1548.93</td>
<td>7</td>
<td>221.28</td>
</tr>
<tr>
<td>42225.69</td>
<td>491</td>
<td>52401.72</td>
<td>52401.72</td>
<td>7</td>
<td>221.28</td>
</tr>
<tr>
<td>86.0</td>
<td>86.0</td>
<td>113.18</td>
<td>113.18</td>
<td>1.96</td>
<td>1.96</td>
</tr>
<tr>
<td>1.06</td>
<td>1.06</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>1.04</td>
<td>1.04</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

### TRIAL VII

<table>
<thead>
<tr>
<th>T</th>
<th>N</th>
<th>S D</th>
<th>Mean</th>
<th>S D</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.7</td>
<td>12.1</td>
<td>44</td>
<td>49.1</td>
<td>9.7</td>
<td>63</td>
</tr>
<tr>
<td>51.6</td>
<td>8.3</td>
<td>65</td>
<td>51.0</td>
<td>7.8</td>
<td>56</td>
</tr>
<tr>
<td>50.0</td>
<td>10.2</td>
<td>109</td>
<td>50.0</td>
<td>9.0</td>
<td>119</td>
</tr>
<tr>
<td>45.9</td>
<td>11.7</td>
<td>57</td>
<td>50.8</td>
<td>11.0</td>
<td>53</td>
</tr>
<tr>
<td>48.3</td>
<td>11.6</td>
<td>110</td>
<td>50.6</td>
<td>11.5</td>
<td>120</td>
</tr>
<tr>
<td>49.6</td>
<td>11.9</td>
<td>55</td>
<td>51.3</td>
<td>11.1</td>
<td>65</td>
</tr>
<tr>
<td>50.6</td>
<td>11.5</td>
<td>120</td>
<td>50.9</td>
<td>9.8</td>
<td>57</td>
</tr>
<tr>
<td>49.8</td>
<td>9.4</td>
<td>75</td>
<td>50.3</td>
<td>9.6</td>
<td>132</td>
</tr>
</tbody>
</table>

### Trials

- **TRIAL IV**: 3, 5, 1, 5, 5, 4, 5, 0, 9, 1, 7, 0
- **TRIAL V**: 5, 5, 4, 5, 0, 9, 1, 7, 0
- **TRIAL VI**: 5, 5, 4, 5, 0, 9, 1, 7, 0
- **TRIAL VII**: 5, 5, 4, 5, 0, 9, 1, 7, 0

### Calculations

- **TRIAL IV**: Mean = 50.0, SD = 10.2
- **TRIAL V**: Mean = 50.0, SD = 9.8
- **TRIAL VI**: Mean = 50.0, SD = 10.0
- **TRIAL VII**: Mean = 50.0, SD = 10.2

---

test, all trials.
Table 8.1  Experimental Treatment and performance, all trials.
Fig. 8.1. Experimental Treatment and performance, all trials.
Fig. 8.2 Experimental Treatment and performance, boys and girls, within treatment, all trials.
Reversal effect

The data from Trials I and II were used in a replication of procedures followed in previous work. A Product-moment correlation coefficient was calculated between the mean pre-test (Trial I) score of a 'matched-pair' (individuals receiving opposite treatments within the same class cell of four) and the difference in gain scores (Trial I to II) of the individuals in the matched pair. These correlations were not statistically significant, and are presented in Table 8.2.

<table>
<thead>
<tr>
<th></th>
<th>N.</th>
<th>r.</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched-pairs</td>
<td>144</td>
<td>-0.0603</td>
<td>NS.</td>
</tr>
<tr>
<td>Matched-pair girls</td>
<td>56</td>
<td>0.0026</td>
<td>NS.</td>
</tr>
<tr>
<td>Matched-pair boys</td>
<td>40</td>
<td>0.1838</td>
<td>NS.</td>
</tr>
</tbody>
</table>

These findings fail to confirm observations of a correlation of 0.2753 reported in previous work.
The mean S.A.I. for each experimental treatment group, (and for boys and girls separately within treatment) were computed for each trial, in the same way as the performance measures reported in section A. The mean S.A.I. scores, and appropriate analyses of variance are presented in the same format as adopted in section A.

The Table shows that the null hypothesis concerning differences between experimental groups can be accepted on trials I and VI. The other trials show differences significant at the .05 and .01 levels. (The .01 levels occur at trials III and IV, where significant differences in performance were found).

Where differences between groups selected by sex within experimental treatment are concerned, the null hypothesis can be rejected over all trials at levels of confidence considerably better than 99%.
Table 8.3
Experimental treatment and SA I, All trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>SAI M</th>
<th>SAI S.D.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boys (+) girls</strong></td>
<td>2.94</td>
<td>3.72</td>
<td>80</td>
</tr>
<tr>
<td>All</td>
<td>1.63</td>
<td>3.59</td>
<td>193</td>
</tr>
<tr>
<td><strong>- (-) girls</strong></td>
<td>2.64</td>
<td>3.73</td>
<td>99</td>
</tr>
<tr>
<td>All</td>
<td>1.97</td>
<td>3.58</td>
<td>197</td>
</tr>
<tr>
<td><strong>Boys (=) girls</strong></td>
<td>3.00</td>
<td>3.30</td>
<td>86</td>
</tr>
<tr>
<td>All</td>
<td>1.90</td>
<td>3.42</td>
<td>190</td>
</tr>
<tr>
<td><strong>Boys (=) girls</strong></td>
<td>2.69</td>
<td>3.32</td>
<td>86</td>
</tr>
<tr>
<td>All</td>
<td>1.78</td>
<td>3.91</td>
<td>189</td>
</tr>
</tbody>
</table>

For all

<table>
<thead>
<tr>
<th>Trial</th>
<th>SS between</th>
<th>df</th>
<th>MS between</th>
<th>SS within</th>
<th>df</th>
<th>MS (with)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boys (+) girls</strong></td>
<td>3.17</td>
<td>3.66</td>
<td>77</td>
<td>2.15</td>
<td>3.52</td>
<td>189</td>
</tr>
<tr>
<td>All</td>
<td>3.06</td>
<td>3.91</td>
<td>92</td>
<td>0.26</td>
<td>3.82</td>
<td>92</td>
</tr>
<tr>
<td><strong>Boys (=) girls</strong></td>
<td>2.18</td>
<td>3.16</td>
<td>84</td>
<td>2.18</td>
<td>4.03</td>
<td>96</td>
</tr>
<tr>
<td>All</td>
<td>1.22</td>
<td>4.18</td>
<td>187</td>
<td>0.94</td>
<td>4.16</td>
<td></td>
</tr>
</tbody>
</table>

For boys and girls groups

<table>
<thead>
<tr>
<th>Trial</th>
<th>SS between</th>
<th>df</th>
<th>MS between</th>
<th>SS within</th>
<th>df</th>
<th>MS within</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boys (+) girls</strong></td>
<td>651.83</td>
<td>7</td>
<td>93.12</td>
<td>9623.73</td>
<td>761</td>
<td>12.65</td>
</tr>
<tr>
<td>All</td>
<td>810.92</td>
<td>7</td>
<td>115.85</td>
<td>9736.11</td>
<td>736</td>
<td>13.23</td>
</tr>
<tr>
<td><strong>Boys (=) girls</strong></td>
<td>651.83</td>
<td>7</td>
<td>93.12</td>
<td>9623.73</td>
<td>761</td>
<td>12.65</td>
</tr>
<tr>
<td>All</td>
<td>810.92</td>
<td>7</td>
<td>115.85</td>
<td>9736.11</td>
<td>736</td>
<td>13.23</td>
</tr>
</tbody>
</table>

**F** values:
- **TRIAL I**: F = 130.79, df = 3, MS = 43.60, P < .05
- **TRIAL II**: F = 326.44, df = 3, MS = 108.81, P < .01
- **TRIAL III**: F = 732.64, df = 7, MS = 104.66, P < .01
- **For boys and girls groups**: F = 1047.7, df = 7, MS = 149.9, P < .01
<table>
<thead>
<tr>
<th></th>
<th>TRIAL IV</th>
<th>TRIAL V</th>
<th>TRIAL VI</th>
<th>TRIAL VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>S.D</td>
<td>M</td>
<td>S.D</td>
</tr>
<tr>
<td>79</td>
<td>2.95</td>
<td>3.31</td>
<td>0.08</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>1.09</td>
<td>3.24</td>
<td>1.08</td>
<td>3.90</td>
</tr>
<tr>
<td>107</td>
<td>1.87</td>
<td>3.39</td>
<td>2.74</td>
<td>4.66</td>
</tr>
<tr>
<td>186</td>
<td>2.05</td>
<td>4.09</td>
<td>0.49</td>
<td>4.52</td>
</tr>
<tr>
<td>96</td>
<td>-0.18</td>
<td>3.94</td>
<td>1.09</td>
<td>4.86</td>
</tr>
<tr>
<td>187</td>
<td>0.94</td>
<td>4.16</td>
<td>0.99</td>
<td>3.76</td>
</tr>
<tr>
<td>85</td>
<td>0.24</td>
<td>3.68</td>
<td>0.34</td>
<td>4.00</td>
</tr>
<tr>
<td>103</td>
<td>-0.09</td>
<td>4.10</td>
<td>0.26</td>
<td>3.93</td>
</tr>
<tr>
<td>188</td>
<td>0.06</td>
<td>3.91</td>
<td>2.89</td>
<td>3.35</td>
</tr>
<tr>
<td>82</td>
<td>3.14</td>
<td>3.63</td>
<td>0.31</td>
<td>4.46</td>
</tr>
<tr>
<td>98</td>
<td>1.18</td>
<td>3.83</td>
<td>1.51</td>
<td>4.18</td>
</tr>
<tr>
<td>180</td>
<td>2.08</td>
<td>3.86</td>
<td>2.80</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;.01</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRIAL IV</th>
<th>TRIAL V</th>
<th>TRIAL VI</th>
<th>TRIAL VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>S.D</td>
<td>M</td>
<td>S.D</td>
</tr>
<tr>
<td>79</td>
<td>2.51</td>
<td>3.86</td>
<td>1.43</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>3.63</td>
<td>0.35</td>
<td>3.60</td>
</tr>
<tr>
<td>107</td>
<td>1.08</td>
<td>3.90</td>
<td>0.82</td>
<td>3.76</td>
</tr>
<tr>
<td>186</td>
<td>2.74</td>
<td>4.66</td>
<td>2.21</td>
<td>3.94</td>
</tr>
<tr>
<td>96</td>
<td>-0.49</td>
<td>4.52</td>
<td>0.28</td>
<td>4.48</td>
</tr>
<tr>
<td>187</td>
<td>1.09</td>
<td>4.86</td>
<td>1.28</td>
<td>4.31</td>
</tr>
<tr>
<td>85</td>
<td>0.99</td>
<td>3.76</td>
<td>1.02</td>
<td>3.46</td>
</tr>
<tr>
<td>103</td>
<td>-0.34</td>
<td>4.00</td>
<td>0.69</td>
<td>3.69</td>
</tr>
<tr>
<td>188</td>
<td>0.26</td>
<td>3.93</td>
<td>0.84</td>
<td>3.59</td>
</tr>
<tr>
<td>82</td>
<td>2.89</td>
<td>3.35</td>
<td>2.44</td>
<td>3.40</td>
</tr>
<tr>
<td>98</td>
<td>0.31</td>
<td>4.46</td>
<td>1.07</td>
<td>3.58</td>
</tr>
<tr>
<td>180</td>
<td>1.51</td>
<td>4.18</td>
<td>1.69</td>
<td>3.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt;.01</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRIAL IV</th>
<th>TRIAL V</th>
<th>TRIAL VI</th>
<th>TRIAL VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>S.D</td>
<td>M</td>
<td>S.D</td>
</tr>
<tr>
<td>79</td>
<td>1.65</td>
<td>4.11</td>
<td>1.65</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td>-0.48</td>
<td>4.06</td>
<td>-0.48</td>
<td>4.06</td>
</tr>
<tr>
<td>107</td>
<td>0.42</td>
<td>4.21</td>
<td>0.42</td>
<td>4.21</td>
</tr>
<tr>
<td>186</td>
<td>2.48</td>
<td>4.29</td>
<td>2.48</td>
<td>4.29</td>
</tr>
<tr>
<td>96</td>
<td>0.52</td>
<td>3.30</td>
<td>0.52</td>
<td>3.30</td>
</tr>
<tr>
<td>187</td>
<td>1.55</td>
<td>3.97</td>
<td>1.55</td>
<td>3.97</td>
</tr>
<tr>
<td>85</td>
<td>1.02</td>
<td>4.00</td>
<td>1.02</td>
<td>4.00</td>
</tr>
<tr>
<td>103</td>
<td>0.76</td>
<td>4.05</td>
<td>0.76</td>
<td>4.05</td>
</tr>
<tr>
<td>188</td>
<td>0.88</td>
<td>4.02</td>
<td>0.88</td>
<td>4.02</td>
</tr>
<tr>
<td>82</td>
<td>2.23</td>
<td>3.42</td>
<td>2.23</td>
<td>3.42</td>
</tr>
<tr>
<td>98</td>
<td>0.68</td>
<td>3.84</td>
<td>0.68</td>
<td>3.84</td>
</tr>
<tr>
<td>180</td>
<td>1.38</td>
<td>3.73</td>
<td>1.38</td>
<td>3.73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRIAL IV</th>
<th>TRIAL V</th>
<th>TRIAL VI</th>
<th>TRIAL VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>S.D</td>
<td>M</td>
<td>S.D</td>
</tr>
<tr>
<td>79</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>186</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>188</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRIAL IV</th>
<th>TRIAL V</th>
<th>TRIAL VI</th>
<th>TRIAL VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>S.D</td>
<td>M</td>
<td>S.D</td>
</tr>
<tr>
<td>79</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>186</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>188</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TRIAL IV</th>
<th>TRIAL V</th>
<th>TRIAL VI</th>
<th>TRIAL VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
<td>SAI</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>S.D</td>
<td>M</td>
<td>S.D</td>
</tr>
<tr>
<td>79</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>186</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>188</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>0.01</td>
<td>&lt;.05</td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 8.3 Experimental treatment and SAI. All trials
Fig. 8.4 Experimental treatment and SAI, boys and girls within treatment, all trials.
SECTION C. Self Assessment Index and Performance

Product-Moment correlations between SAI and T scores within each experimental treatment group on Trial I were carried out and are presented in table 8.4.

Table 8.4. Correlations between SAI and performance within treatments. Trial I.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys +</td>
<td>70</td>
<td>.41</td>
</tr>
<tr>
<td>Girls +</td>
<td>105</td>
<td>.35</td>
</tr>
<tr>
<td>Total +</td>
<td>175</td>
<td>.40</td>
</tr>
<tr>
<td>Boys -</td>
<td>90</td>
<td>.30</td>
</tr>
<tr>
<td>Girls -</td>
<td>90</td>
<td>.47</td>
</tr>
<tr>
<td>Total -</td>
<td>180</td>
<td>.40</td>
</tr>
<tr>
<td>Boys *</td>
<td>79</td>
<td>.43</td>
</tr>
<tr>
<td>Girls *</td>
<td>100</td>
<td>.26</td>
</tr>
<tr>
<td>Total *</td>
<td>179</td>
<td>.40</td>
</tr>
<tr>
<td>Boys =</td>
<td>80</td>
<td>.36</td>
</tr>
<tr>
<td>Girls =</td>
<td>99</td>
<td>.47</td>
</tr>
<tr>
<td>Total =</td>
<td>179</td>
<td>.42</td>
</tr>
</tbody>
</table>

It should be noted that no experimental treatments had been applied at trial I. The groups therefore represent random samples from within the experimental population.
SAI and Ability

The mean SAI and Standard deviation of boys and girls within crude ability groupings on Trial I are shown in Table 8.4. below.

Table 8.4. Mean SAI, SD within ability levels

<table>
<thead>
<tr>
<th></th>
<th>SAI</th>
<th>SAI</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>&quot;Failing&quot; boys</td>
<td>3.63</td>
<td>3.47</td>
<td>171</td>
</tr>
<tr>
<td>&quot;Failing &quot; girls</td>
<td>2.94</td>
<td>3.38</td>
<td>110</td>
</tr>
<tr>
<td>&quot;Average&quot; boys</td>
<td>2.67</td>
<td>3.14</td>
<td>115</td>
</tr>
<tr>
<td>&quot;Average&quot; girls</td>
<td>1.23</td>
<td>3.13</td>
<td>128</td>
</tr>
<tr>
<td>&quot;Successful&quot; boys</td>
<td>1.23</td>
<td>3.40</td>
<td>89</td>
</tr>
<tr>
<td>&quot;Successful&quot; girls</td>
<td>-0.56</td>
<td>3.45</td>
<td>156</td>
</tr>
</tbody>
</table>

The "Failing", "Average" and "Successful" ability levels for table 8.4 above, represent approximate thirds of the ability distribution based on equivalent percentile rank of T. Score on Trial I. Thus "Failing" pupils represent the bottom third of the whole sample, "Average" pupils the middle third, and "Successful" the upper third.
CHAPTER 9

THE MAIN STUDY RESULTS: DISCUSSION

Introduction

The results of the main study fall into three major categories which are represented in this chapter as three sections;

Section A; concerned with findings regarding experimental treatment and performance.

Section B; concerned with findings regarding experimental treatment and self-assessment.

and Section C; concerned with findings regarding the relationship between self-assessment and performance.

The seven tests of the experiment may be further classified into two sub-experiments. The first sub-experiment (trials I to IV) permits comparison between treatment effects upon groups of equivalent ability and equivalent pre-experimental history of feedback (within the limits of random sampling). The second sub-experiment (trials IV to VII) permits comparisons between four equivalent ability groups each of which has been exposed to a different history of 'pre-experimental' feedback. The comparisons in the second sub-experiment are therefore comparisons of four differing sequences of treatment.

The findings in each of the above sections will be discussed both in general terms and with respect to any differences between the sexes.
SECTION A: Experimental treatment and Performance

A.1. Treatment comparisons, (trials I to IV) general (Fig.8.1)

No significant differences in performance were found on the pretest (trial I) or following the first application of experimental treatment (trial:II). Significant differences appear following the second, and third applications of experimental treatment where the superior performance of the 'actual' group over the 'artificial' groups is most evident. The 'actual' group shows a consistent improvement following all three applications of experimental treatment, in contrast to the other three groups. The artificial discrepancy groups ('positive' and 'negative') show trends of deterioration and improvement following successive applications of their respective treatments and these trends are similar. The 'matching' group displays a trend opposite to that occurring in the artificial discrepancy groups, and shows the greatest relative improvement from trial to trial and the greatest relative deterioration from trial to trial (over trials II to IV) of all the artificial treatments. (see fig. 8.1.)

A.2. Treatment comparisons, (Trials I to IV) by sex. (Fig.8.2.)

Statistically significant differences were found between boys and girls groups on trials I to III, showing initial (pre-test) differences which increased in statistical significance up to the second application of experimental treatment, becoming non-significant thereafter. The only consistent trend in the data was shown by girls receiving 'actual' treatment, who improved following all three applications of experimental treatment. Boys exposed to 'actual' treatment were less consistent, but showed the
greatest improvement of all the experiment groups (when separated by sex) over all three applications of treatment.

No consistent trends of improvement and deterioration were observed in the artificial treatment groups, and indeed, boys and girls responded differently to successive applications of the same treatment. Perhaps the most striking effects were as follows:

(i) **Positive Treatment.**
Boys show a consistent slight improvement following one and two applications of treatment, whereas girls deteriorate initially, and almost recover following a second application. Both groups deteriorate following a third application.

(ii) **Negative Treatment.**
Girls deteriorate after the first application of treatment, whereas boys improve slightly. Girls then improve considerably following the second application (ranking top on trial III) while boys deteriorate; and both groups deteriorate following a third application, the girls showing considerably greater deterioration.

(iii) **Matching Treatment.**
Girls deteriorate consistently following one and two applications of experimental treatment, while boys show a slight improvement following the first application, and a considerable deterioration following the second application. Both groups improve following the third application.
A.3. Treatment comparisons: discussion.

The results of the first sub-experiment within the main study would seem to indicate that there is no advantage to be gained in the employment of artificial distortion of marks for motivational purposes. When artificial discrepancies of 10% of the range of test scores are returned to individuals, the effect upon performance is not consistent over a number of applications and not consistent in effect between boys and girls at any point in a series of applications. Further, all mean performances (except girls, negative trial III) following any artificial treatment; including the return of the subjects own expectation, are inferior to performances following the return of an actual mark. The reasons for the superiority of actual mark will be discussed further in Chapter 10.

A.4. Sequence comparisons (Trials IV to VII) (Fig. 8.1)

For the following discussion, Trial IV may now be regarded as a pre-test score for four equivalent groups exposed to four differing 'histories' of experimental feedback. No statistically significant differences in performance were found at any trial (except IV) constituting the second sub-experiment within the main study. This is itself an interesting finding. The statistically significant differences between the groups performances on trials III and IV disappear following a change of experimental treatment to an 'opposite' form of treatment. Inspection of fig. 8.1 shows that the mean performance scores of the experimental groups are distributed over approximately the same range on trial V as on trial I, though the relative rank position of the groups is different. It is reasonable to infer from this finding that a reversal of experimental treatment can produce
a reversal of performance trends, though the case of the "matching followed by actual" group is different to interpret. Perhaps the most interesting trends are those shown by the "actual followed by matching" group (a considerable deterioration following change of experimental treatment) and by the "negative followed by positive" group (a slight improvement followed by considerable deterioration after change of experimental treatment). A comparison between the degree of improvement (Trial IV to V) of the opposite groups, "negative followed by positive" and "positive followed by negative" is also of interest; the former group showing the largest performance increment and the latter the smallest performance increment.

A.5. Sequence comparisons (Trials IV to VII) by sex (Fig.8.2).

When performance differences between the sexes are examined, the picture is as confused as in the first sub-experiment. Boys and Girls seem to respond similarly to a change from 'actual' to 'matching', and to a change from 'positive' to 'negative'; at least following the first application of the new treatment. Thereafter, no consistent trends are apparent.

The differences between boys and girls groups approach statistical significance on trial VII, where the "negative followed by positive" boys group show extraordinary deterioration following the second application of opposite treatment.

When the effects of sequence of the artificial discrepancies are studied by sex, and the data from all trials following a particular treatment are averaged, the difference between the performance of the sexes are perhaps more clear;
Table 9.1. Mean T scores (3 trials) by sex, positive and negative treatment. Gain, First to Second treatment

<table>
<thead>
<tr>
<th></th>
<th>Positive first.</th>
<th>Positive second: Gain 1st to 2nd.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Boys.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.4 (N190)</td>
<td>47.9 (N189),</td>
<td>gain = -0.95</td>
</tr>
<tr>
<td>49.6 (N279)</td>
<td>50.93 (N208),</td>
<td>gain = +2.22</td>
</tr>
<tr>
<td><strong>Girls.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.85 (N249)</td>
<td>44.8 (N143),</td>
<td>gain = -3.60</td>
</tr>
<tr>
<td>48.71 (N232)</td>
<td>51.15 (N208),</td>
<td>gain = +1.55</td>
</tr>
</tbody>
</table>

It would appear, from Table 9.1, that boys and girls performances are relatively unaffected by sequence as far as positive treatment is concerned (both groups having similar mean scores whether positive treatment came first or second in sequence) but considerable difference in sequence effect under negative treatment. Here, boys appear to deteriorate considerably and girls appear to improve, when negative treatment is second in sequence. Boys and girls scores (when negative treatment is applied first) are almost identical. Boys would appear to respond badly to a sequence of receiving 'generous' marks and then 'severe' marks, whereas girls improve under this sequence. It should be noted that girls improve still more when this sequence is reversed, and that boys deteriorate less.

A.6. The 'reversal' effect.

As noted in earlier chapters, the previous work had indicated that positive and negative treatments appeared to reverse incentive effects at differing ability levels, so that negative treatment was a better incentive at high ability.
levels than positive treatment, and that these effects reversed at lower ability levels. The main study finding is that this effect is not present on trials I and II. It should be noted that the sample studied in previous work (Pickup and Anthony, 1968) was one for which regular classroom tests were a usual feature of school experience in the school subjects included. The sample for the present main study included only a small proportion of pupils who were used to regular classroom tests, and a large proportion of pupils who were not regularly exposed to testing. It is considered that this is a likely explanation of failure to replicate the findings of previous work, and the subject will be discussed further in chapter 10.
SECTION B: Experimental Treatment and Self-Assessment

B.1. Treatment Comparisons (Trials I to IV) (Fig. 8.3.)

Significant differences in S.A.I. are found following the first application of experimental treatment, and the significance of the differences increases after each successive application of treatment. The trends are much more consistent than those observed of performance, and can be seen to fall into two types;

(i) A consistent progressive decline in S.A.I. is a feature of both the 'actual' and the 'negative' treatments, with actual treatment showing the greatest decline.

(ii) An initial increase in S.A.I. followed by a slight decline is a feature of both 'matching' and 'positive' treatment.

All groups display an initial mean S.A.I. of some 1.8 marks, which is reduced to almost zero after three applications of 'actual' treatment and approximately halved after three applications of 'negative' treatment. Thus, S.A.I. can be reduced by the application of fixed-discrepancy negative treatment, or by the application of variable-discrepancy negative treatment ('actual' treatment).

S.A.I. is initially increased, and remains larger than at pre-test level, following application of a series of 'matching' or 'positive' treatments.

B.2. Treatment comparisons (Trials I to IV) by sex. (Fig. 8.4.)

Statistically significant differences between boys and girls groups within treatment are evident at every trial during the entire series. The findings are; that boys overestimate
their performance score to a much greater extent than girls, except after two applications of 'actual' treatment, when their S.A.I. is lower than that displayed by girls receiving 'positive' and 'matching' treatments. The boys and girls groups within a particular treatment show similar trends over successive applications, but the only groups which underestimate their performance at any point are girls exposed to 'actual' and 'negative' treatments.

B.3. Sequence Comparisons (Trials IV to VII) (Fig. 8.3.)

Statistically significant differences between treatment groups are found at the pre-test (trial IV) and are increased following the first application of the second type of treatment in the sequence. There is no statistically significant difference after the second application of the second treatment, and significant differences reappear after the third application. Again, the general trends for the actual and negative treatments (second in series) are similar, as are the trends for the positive and negative treatments. It would appear that the 'positive followed by negative' treatment group shows a greater decline in S.A.I. than the 'matching followed by actual' group; in other words, that actual treatment is less effective in reducing S.A.I. when it is preceded by a history of matching treatment.

When matching and positive treatments are applied following 'opposite' histories, the increase in S.A.I. is more consistent than shown in the first sub-experiment (section B.1.) where these treatments are first in sequence.
B.4. Sequence Comparisons (Tests IV to VII) by sex. (Fig. 8.4.)

Statistically significant differences are found between boys and girls groups within treatment at all trials, the lowest significance level being that at trial IV. Again, broadly similar trends by each sex within the two treatments 'actual' and 'negative', and the two treatments 'matching' and 'positive' are shown, but the different trends between sexes are of interest. Girls appear to maintain a trend established in the first four trials, changing direction of trend only after the second application of the new treatment, when all girls groups increase in S.A.I. This increase is maintained following the third application of treatment in the 'matching' and 'positive' groups, while the 'actual' and negative groups reduce S.A.I. In contrast, boys groups exposed to 'matching' and 'positive' treatments for the first time following an opposite history show immediate increases of S.A.I, which are later maintained. The greatest difference in gain of S.A.I. (from trial IV to V) is shown between boys exposed to either positive or negative treatment as a second treatment in a series.
SECTION C: Self Assessment and Performance. (table 8.3.)

The observation of eight random groups of boys and girls on trial I (pretest) reflects the relationship between S.A.I. and performance in a pre-experimental context. The average correlation of -0.405 is therefore an indication of relationships to be found in normal classrooms. Table 8.4 shows the mean size of S.A.I. for differing ability groups and demonstrates the negative correlation between ability and S.A.I. The negative correlation between S.A.I. and ability is confounded to some extent by the constrained nature of the S.A.I. employed in this study. It could be argued that the probability of over-estimating a particular performance is reduced at the higher levels of ability, and that this fact alone could account for a negative correlation.

However it is of interest to note that boys mean S.A.I. is higher than girls mean S.A.I. at all ability levels; and while the effects of a constrained S.A.I. can be theoretically demonstrated at the extremes of ability (note the differing numbers of boys and girls in the "Failing" and "Successful" thirds of the total sample) these effects can not be held to apply to the difference between the sexes in the "Average" third (where the numbers of boys and girls are approximately equal). It would appear, therefore, that differences in S.A.I. can occur for reasons other than mere performance level, and that the statistical effects of constraining S.A.I. are of minor importance in the interpretation of the observed negative relationship between S.A.I. and performance.
PART III
CONCLUSIONS AND IMPLICATIONS
CHAPTER 10

CONCLUSIONS AND IMPLICATIONS

Introduction

This chapter presents the final conclusions of the present study, comprising the work reported in Parts I and II, the "Laboratory" and "Main" studies: and where possible, relates this work to comparable studies carried out by other investigators.

The first section of this chapter is concerned with the effects upon performance of various forms of feedback. The second section is concerned with observations of self-assessment following performances; and the chapter concludes with a discussion of implications for further research and for classroom practice.

A. FEEDBACK AND PERFORMANCE

Investigations in this area have been concerned fundamentally with the motivational, or incentive effects of those forms of feedback which may induce positive, negative, or neutral 'affect' in the recipient. Experiments have been conducted in a range of environments which may be broadly classified into 'laboratory' and 'field' experiments. The present study has findings derived from both of these types of investigation and these findings will be discussed in the appropriate sub-sections which follow.

A1 Laboratory investigation

While none of the findings of the laboratory study reported in part I reached an acceptable level of significance, certain trends in the data indicated that negative treatment (feedback which could be assumed to
induce negative affect in the recipient) was a superior incentive for improved performance than either positive or neutral treatment. Among the deficiencies of the design were; a small group size (N=10), a small discrepancy (10%), and failure to control for task interest in the subjects. No other study has been found which investigates positive or negative feedback relative to the subjects' self-assessment, though two other studies have a related field of observation, and contain data which may be used for comparison.

Zajonc and Brickman (1969) in a complex laboratory investigation of 'aspiration'-type cognitive phenomena (termed 'expectancy' by the authors) showed that negative feedback (indication of failure) proved a significant incentive to improvement of reaction-time (in a stimulus-response situation); and that negative feedback was a superior incentive to positive feedback; both being superior to no feedback. These findings constituted a minor area of interest to the authors and are not taken into account in their conclusions. Indeed, at one point in the text the authors contradict their findings by attributing the greatest improvement to the groups exposed to positive feedback. This error is of interest in itself, as will be shown.

Zajonc and Brickman then embarked upon an investigation of possible asymptotic performance levels prior to feedback application. They considered that some
subjects may have reached a ceiling of potential improvement in the four trials preceding the application of feedback. They do not report whether this was an attempt to explain the superiority of the negative groups, but state that an analysis of covariance, with pre-feedback improvement as the covariate "... did not substantially change the results ..."

From the findings in their report it is possible to 'correct' for improvement prior to feedback, and to establish the variation for which each feedback type is responsible. When these calculations were performed it was found that positive treatment was responsible for only 63% of the improvement following feedback of that treatment group; the remainder being associated with improving trends prior to feedback. Where negative feedback is concerned Zajonc and Brickman's findings indicate that negative treatment is most effective when applied to those who have attained a 'ceiling' of prior improvement, and that overall, almost twice as much improvement can be attributed to negative, as opposed to positive treatment.

Snyder (1972) investigating comparison level feedback effects upon classroom-related verbal learning performance concludes that subjects who were informed of a low comparison level (a standard achieved at the seventeenth percentile by previous subjects) performed
significantly better than subjects informed of a high comparison level (standard achieved at the eighty third percentile level by previous subjects) who in turn performed significantly better than students receiving no comparison-level feedback. The similarity of these forms of feedback to the 'social' treatments investigated in the laboratory study reported in part I is evident. Snyder's 'high comparison level' subjects are being informed of a high standard of previous performance, and may be compared with subjects in the present study who were informed that their performance was 'below average': a negative social treatment.

Each subject was informed of the appropriate comparison level at the top of the answer sheet for each trial but Snyder's report does not make clear how this was to be considered as application of experimental treatment; except that the main research interest was to be effects upon performance. Presumably effects upon attention during the learning task prior to performance measure are confounded with performance effects for the latter two trials in the series of three. Examination of the trends in Snyder's data show that performance differences occur on the first trial, the low comparison level subjects (positive treatment) being superior in performance to the other two groups. When all three trials are considered, the low comparison
(positive) and no comparison (non-social) groups show the same improvement trends, maintaining the difference on trials two and three which was apparent on trial one. The high comparison level group (negative social treatment) however, showed a relatively greater improvement over trials than did the other two groups, and Snyder reports a significant interaction effect between comparison level received and trials, attributing this interaction effect to the performance over trials of the high treatment group. This finding is clearly inconsistent with Snyder's conclusions and it is surprising that no attempt was reported to test for initial differences in ability between treatment groups, complete reliance being placed upon randomisation of subjects to experimental treatment. Snyder's conclusions are somewhat erroneous when his findings are examined in detail; it is difficult to reconcile his statement: "The results of the present and previous studies indicate the consistency of achieving improved performance through the low comparison level" with the published findings, or indeed an earlier statement: "The detrimental effect on the performance of the high comparison level, therefore, occurs initially in the experiment and diminishes over the trials". Since no evidence was advanced to show that these treatment groups were indeed equal in ability at the start of the experiment, the effects noted by Snyder may be due to initial differences in ability, and on his own admission the
high comparison level (low treatment) group shows the greatest gain in performance (from trial 1) over successive applications of treatment.

It would appear that evidence supporting the findings of the laboratory study exists in the two investigations here considered: it would appear that, in certain circumstances it is possible for negative treatment to have superior incentive value to positive treatment. It is interesting to note that this type of finding has not been recognised, or not been given prominence in the conclusions of the workers here quoted. A similar lack of recognition was found in the Leith and Davis (1969) research reported in the next sub-section.

A2 Field Investigation

The findings of the main study (reported in part II) with respect to positive and negative treatment do not initially appear to support the conclusions stated in the previous sub-section. When findings related to change of experimental treatment from a previous controlled history are examined, however, these findings do support the view that negative treatment can prove a superior incentive to positive treatment. The findings of the first sub-experiment (initial applications of treatment) do not demonstrate important differences between groups exposed to fixed, artificial discrepancies of 10% either above or below the subject's self-assessment. The findings do demonstrate the
superiority of actual treatment (receiving an undistorted external assessment) over all other treatments. It can be shown that the majority of subjects exposed to 'actual' treatment are, in fact being exposed to a form of negative treatment. The findings related to Self-Assessment Index (discussed later in this chapter) show that a large proportion of subjects in the 'actual' treatment groups are being presented with negative information considerably larger than the 10% discrepancy adopted for the formally classified 'negative' treatment. The superiority of actual treatment as an incentive must therefore be considered as being partially dependent upon negative feedback with a range of discrepancy size far beyond that employed in the fixed discrepancy negative treatment.

Leith and Davis (1969) investigating the effects of verbal incentives in a programmed learning task conclude, from their four-class (N=120) eight-test experiment that "This study confirms the trend identified in eleven recent experiments by Kennedy and Willcutt (1964) that praise improves performance while reproof has the opposite effect". Their report does not indicate how this effect was established, or indeed the statistical significance of any differences. Their graph of mean performance over all eight trials does, however, indicate that, while the 'positive' and 'neutral' treatment groups
show a progressive decline in scores (indicating progressive difficulty of the tests) the 'negative' treatment group (reproof) maintains its score after two applications of treatment for a further two trials, indicating greater effort than the other two groups for this period. This effect was not noted by the authors.

Means and Means (1971) investigated the effects upon achievement of prior information concerning aptitude. They found that students who had previously attained a high grade point average performed better with negative information concerning aptitude; while previous low grade point average students performed better with positive information.

Pickup and Anthony (1968) found that the effectiveness of positive and negative forms of feedback as incentives reversed as the opposing treatments were applied at different levels of pre-test score. The present main study contained procedures which replicated the conditions of this finding, and which failed to confirm that such a 'reversal effect' exists. (See para. A5, chapter 9). As noted in chapter nine, there were important differences between the samples studied in the 1968 report, and the samples here studied, with respect to pre-experimental history. When the findings of the second half of the present main study are examined (the effects of sequence of experimental treatment) some support for the existence of this effect
can be found. As noted earlier, negative treatment following a history of positive treatment produces an immediate improvement in performance, as does positive treatment following a history of negative treatment (though to a lesser extent). Both the present study and the 1968 study demonstrate that the Self-Assessment Index of low ability pupils is considerably higher than that of high ability pupils, and that indeed, many high ability pupils have a negative Self-Assessment Index. It follows that, in a context of regular classroom tests (a feature of the 1968 sample) the high ability subjects have been exposed to a history of positive treatment, and the low-ability subjects have been exposed to a history of negative treatment. The main study findings indicate that these groups would improve when exposed to one application of opposite treatment: thus the 1968 high ability subjects would improve following negative treatment, and the low ability subjects would improve following positive treatment. This is essentially what was found in the 1968 study, though it is now apparent that the effect was less dependent upon ability than upon ability and history of feedback experience.

B. THE SELF-ASSESSMENT INDEX

The relationship between a subject's self-assessment and an assessment from an external source has been a topic which has been part of previous
investigations. Ringness (1961) in an investigation of reality of self-concept among children of varying intelligence reports that the self-estimate varies with intelligence, sex, and the situation. While all groups studied tended to over-estimate their success in common school tasks, the low IQ groups more generally over-estimated their success than the 'bright' or 'average' groups. Murstein (1965), studying the relationship of grade expectations to actual grades received, found that almost no student perceived himself as a poor or even mediocre student. Further, "Low" (ability) subjects tended to be grossly unrealistic in their initial estimates and to be relatively refractory to the effects of experience.

Pickup and Anthony (1968) found that a large majority of pupils (72.6%) over-estimated their mark on the initial test, and that low test scorers were especially liable to over-estimate their mark. When the discrepancy between 'expected' mark and 'actual' mark was correlated with pre-test score, r was found to be -0.33 (P<.001). This finding is supported by the present study, where the "discrepancy between expected mark and actual mark" is termed the Self-Assessment Index.

As noted in earlier chapters, the present study has not been designed to investigate the phenomenon of Self-Assessment Index, and so can present only initial observations of the variable. It would appear that
consistent negative relationships between performance and Self-Assessment Index are observations of some importance, requiring further investigation. It would also appear that boys and girls differ with respect to this phenomenon, boys over-estimating initially by a mean of 14% of the possible range of scores, and girls over-estimating by a mean of 5%. The standard deviations of both groups are almost identical, at approximately 3.5 (17% of the range of possible test score).

C. IMPLICATIONS

(i) For further research:

It would seem to be important to investigate still further the influence upon future performance of feedback conditions which may be supposed to induce negative affect in individuals in learning contexts, and to establish under what circumstances these forms of feedback can be effective incentives. There appears to be a reluctance to perceive or comment upon incentive effects of these negative forms of feedback which may derive from the expectations of those research workers who have included such experimental treatments in their designs. Certainly, the incentive effects of these forms of feedback are not being given the attention that they would seem to deserve. Further investigation of the phenomenon of self-assessment would appear to be a worthwhile line of enquiry, particularly in
relation to ability or performance. It is interesting to speculate that discrepancies between "interior" and "exterior" assessment may indicate the degree of effort expended on a task, and that the observed relationship between S.A.I. and performance is a result of both variables being dependent upon the motivational state of the individual. Such a hypothesis is consistent with the findings of this study.

(ii) For classroom practice

It would appear that there is no advantage to be gained by artificially distorting marks or grades for motivational purposes, as a group treatment. Such effects as do appear when artificial distortion is employed seem to be inconsistent, or dependent upon specific situational factors. The best policy appears to be an honest return of an external assessment.

The procedure of obtaining a self-assessment from an individual in a test situation would appear to be a useful procedure when the assessor is also involved as a teacher with that individual's future performance or learning. Where an individual's S.A.I. is excessively large, this fact can be noted by the teacher, and would provide a useful indication of the 'unrealistic' criteria being adopted by the individual. It is unlikely that a learner can improve, or reach full potential when the learner's criteria are different from those held by an experienced assessor, though this
places considerable responsibility for accurate and honest judgement upon the assessor.

Hammer (1972) investigating the effects of differential, written teacher comments on student performance, showed that when these comments incorporated the students grade expectation, future test performance was significantly superior to that shown by groups whose expectation had not been incorporated. Hammer suggests that the adoption of these procedures is superior to 'individualising' feedback on an intuitive basis, utilising past experiences and impressions of the student. With this, the present author can only concur.
APPENDICES
APPENDIX A

Previous work; Reliability Indices etc.

<table>
<thead>
<tr>
<th>Class Code</th>
<th>Sex of Group</th>
<th>Subject</th>
<th>Test</th>
<th>Remarks</th>
<th>SD</th>
<th>Reliability Estimate</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>Girls Maths</td>
<td>(a) Arith/Geom.)</td>
<td>equiv.</td>
<td>2.42</td>
<td>.3927</td>
<td>4.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Arith/Geom.)</td>
<td></td>
<td>2.51</td>
<td>.2431</td>
<td>8.26</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>Boys English</td>
<td>(a) Spellings)</td>
<td>equiv.</td>
<td>4.39</td>
<td>.7525</td>
<td>11.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Spellings</td>
<td></td>
<td>4.2</td>
<td>.7655</td>
<td>12.25</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Girls English</td>
<td>(a) Comprehension</td>
<td></td>
<td>3.987</td>
<td>.7113</td>
<td>7.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Spelling</td>
<td></td>
<td>3.619</td>
<td>.6952</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>Girls Maths</td>
<td>(a) Squares, Sq.roots.</td>
<td></td>
<td>3.22</td>
<td>.5491</td>
<td>10.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Algebra</td>
<td></td>
<td>5.6</td>
<td>.8877</td>
<td>8.84</td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>Boys English</td>
<td>(a) Spellings)</td>
<td>equiv.</td>
<td>4.36</td>
<td>.7700</td>
<td>10.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Spellings</td>
<td></td>
<td>3.502</td>
<td>.6325</td>
<td>10.2</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>Boys Maths</td>
<td>(a) Sets</td>
<td></td>
<td>4.03</td>
<td>.8828</td>
<td>10.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Arith. problems</td>
<td></td>
<td>3.45</td>
<td>.6141</td>
<td>10.25</td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>Girls English</td>
<td>(a) Comprehension</td>
<td>equiv.</td>
<td>3.79</td>
<td>.7014</td>
<td>11.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Comprehension</td>
<td></td>
<td>1.92</td>
<td>.3472</td>
<td>11.56</td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>Girls Maths</td>
<td>(a) Algebra</td>
<td>equiv.</td>
<td>4.123</td>
<td>.7430</td>
<td>10.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Algebra</td>
<td></td>
<td>4.44</td>
<td>.7564</td>
<td>10.05</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>Boys English</td>
<td>(a) Spellings</td>
<td></td>
<td>4.025</td>
<td>.7958</td>
<td>13.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Spellings + nouns</td>
<td></td>
<td>4.775</td>
<td>.9401</td>
<td>12.71</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>Boys Maths</td>
<td>(a) Probs, Decimals</td>
<td>equiv.</td>
<td>3.709</td>
<td>.6600</td>
<td>8.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Probs, Decimals</td>
<td></td>
<td>4.719</td>
<td>.8801</td>
<td>9.27</td>
<td></td>
</tr>
</tbody>
</table>

(equiv. = 'equivalent forms' of the same material)
### PERFORMANCE SUMMARY

**NAME** ...................... M/F  
**ED. DIVISION** [ ] **BOOTH** [ ] **GROUP** [ ]

<table>
<thead>
<tr>
<th>TRIAL</th>
<th>EXPECTED SCORE</th>
<th>ACTUAL SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SCORING USE ONLY:**

<table>
<thead>
<tr>
<th>CHECKED</th>
<th>INITIALS</th>
<th>CORRECT/Incorrect</th>
</tr>
</thead>
</table>

ASP/3/3
APPENDIX D

EXPERIMENT IN LEARNING

INSTRUCTIONS: SCORER: */-

1. Select, at random, a card from the box to determine the booth you will be scoring on this run. Return the card.

2. Place 4 copies of score sheet (A.J.P./D/I.) and 1 copy of Performance Summary in your booth.

3. After each trial, bring the checked score form to the control console, wait until subject is informed of his/her score, and collect your form again. At the conclusion of the experiment, check your subject's performance summary against his/her score sheets and hand in all data.

SCORING INSTRUCTIONS

TRIAL 1: Take subjects score sheet when he/she signifies completion. Score all correct responses wherever they appear on the form (This applies to all trials). Bring to control console for check. Keep this, and subsequent forms.

PUM. NOJ. VIC. JUX. QOC. ZAJ. FEC. DOW. CUQ. ROP.

Enter your scoring code in space marked "initials". Subject's score in box "score"

TRIAL 2: Score as in trial 1.

RUF. VIN. ZOM. BEF. XBB. MEP. GIW. CBB. HET. GAH.

Scoring code in "initials" Subject's score in "score" Experimental treatment; On dotted line next to "Booth no" do as follows:

Insert subject's real score. X. Insert "-" here

TRIAL 3: Score as before.

KOM. RIY. PIF. MUB. YIL. GIK. NAR. YUD. QEM. LON.

Experimental treatment; As in trial 2.

TRIAL 4: Score as before.

YBE. FID. TOV. PIL. WUY. KOF. QAS. XAW. VIN. YOF.

No experimental treatment. Scoring code in "initials". Subject's score in "score".

At the conclusion of the experiment, check the subject's Performance Summary against his/her score sheets. Hand in all data - and this leaflet.

Note: If subject expects 0, 1, 9, or 10 as expected mark on trials 2 or 3 write N.T. on dotted line.
"Please do not touch any of the controls on the console in front of you.

This experiment is concerned with the learning of verbal material and with self-evaluation of performance. You will be asked to memorise a list of ten, three-letter 'nonsense syllables' in a period of 20 seconds. There will be four trials of this kind. It is not expected that you will be able to memorise all ten syllables.

On the desk you will find four copies of a form marked AJP/D/1. This is your score sheet. On this score sheet you are asked to fill in your booth number in the space top left of the form. Would you take one of these forms now, and write your booth number in the appropriate space, and, in the space below that insert the number 1. Please do not write your name on this form; or write anything on the dotted line next to the space marked "booth number" or in the spaces designated "For scorers use only".

There are ten blocks on this form; and a space for entering your expected score. You are asked to write in these blocks the nonsense syllables that you have memorised. If you can not remember a particular nonsense syllable you are asked to make an attempt, or your best guess, so that each block on the score sheet contains an attempt.
Please Note it does not matter in which order you write the nonsense syllables. A response if correct will be scored regardless of its position on the score sheet.

The first trial of this experiment will be for practice purposes only; to ensure that you will know the experimental conditions and will have had experience of the type of material and the length of time you have to memorise the material.

The procedure for all the trials will be as follows: I will ask you to place your pens and pencils down and a five-second count-down, that is, 5-4-3-2-1, will be given as a warning that the nonsense syllables are to be shown in the area now illuminated.

(At this point the overhead projector was switched on, illuminating the screen behind the experimenters console)

The group of syllables will be shown for twenty seconds. You should have your pens and pencils on the table throughout this period. Do not attempt to write the syllables down. This is impossible anyway!

When the 20 second period expires, that is, when the screen goes dark.

(At this point the overhead projector was switched off)

You will have two minutes to fill in your score sheet, and some background music will be played for your
entertainment. When you have completed filling in the blank spaces on the form, insert in the space provided, your estimate of the number of nonsense syllables you have correctly memorised.

Now will you please turn your attention to the other form of which you should have one copy headed "Performance Summary".

Would you now write your name and Education division in the appropriate spaces on the form (TEN SECOND PAUSE). Please strike out the letter M if you are female, or the letter F if you are male. Leave the space marked 'group' for the present.

On this form you will see four blocks for each of the four trials. For each trial you are asked to insert your expected score in the square which contains the dotted line. When you have done this raise your hand. Your scorer will then collect your scoresheet.

You will not be informed of your actual score on trial 1, so the space for this has a cross in it. On all subsequent trials you will be informed of your actual score; please insert this in the appropriate space as soon as you are so informed.

Please keep your headphones on throughout the experiment, and please do not communicate in any way with anyone except your scorer."
APPENDIX F.

INFORMATION FOR PARTICIPATING TEACHERS, MAIN STUDY

Classroom Incentives Project

Information for participating teachers

Introduction

This research project is the second stage of a long-term investigation of the incentive value of classroom marking. Previous work has established a number of interesting findings about the effects of teacher's marks and the present study is concerned with confirmation (or otherwise) over a large sample of schoolchildren.

The research, to be valid and useful to teachers, has to be conducted in a manner which closely resembles normal classroom practice. The children should have no knowledge that they are taking part in an experiment - this can only be achieved by the cooperation and involvement of teachers who are prepared to act as research workers. While the research has been designed to minimise disruption of normal school procedure, and to present you with as small an additional load as possible, participation in this project will make an extra demand upon your time. We are very grateful for your support, and will, of course, send you details of the research findings at the conclusion of the project.

Summary of procedures

Each participating teacher is asked to construct and mark seven 20-item tests of an area of their normal classwork. The tests need only be brief: for example (i) 20 spellings, or (ii) 20 Mental arithmetic problems. The tests can be administered over any reasonable period, though weekly tests may be most appropriate; and pupils answers should be written on loose paper.

At the conclusion of each test the participating teacher is asked to distribute additional slips of paper to each pupil upon which the pupil is asked to write (i) his or her name and class; and (ii) the mark that he or she expects to receive for that test.

When the test has been marked, the participating teacher is asked to record marks and expected marks on a form provided: and return a mark on the 'pupil's expectation' slip, retaining the test script.

Description of the research

Previous findings have indicated that discrepancies between teachers marks and pupil's expectations produce incentive effects. The present research is concerned with the experimental manipulation of teachers marks over a short period to further examine these effects.
The experiment consists in returning to pupils a mark which is based upon their expected mark rather than returning their actual score on a test. The procedure of returning marks in this way is the experimental treatment. There are three experimental treatments and a control treatment.

1. 'High' treatment. In this condition pupils are informed that they scored 2 marks more than they expected, regardless of their actual score.

2. 'Low' treatment. Here pupils are informed that they scored 2 marks less than they expected regardless of their actual score.

3. 'Matching' treatment. Pupils are informed that they scored the same as their expectation, regardless of their actual score.

4. 'Control'. In the control condition pupils are returned their actual score.

Each pupil will experience two differing treatments throughout the 7 tests. The experimental treatments, to be effective, should be returned by the teacher prior to the work to be examined in a subsequent test, but where this is not applicable, prior to the subsequent test.

Construction of the tests

Since a number of school subjects, age and ability ranges will be represented in the project, selection of material, length and format of tests is best achieved in discussion with your research contact, but the following points may be valid.

1. Tests should be constructed of easily scored 'objective-type' items which require no subjective interpretation.

2. Each test should be as independent of each other as possible in terms of the material to be tested.

3. Each test should be constructed so as to give each pupil an opportunity of gaining some marks, but extreme difficulty in gaining full marks.

4. The pupils may be given the impression that they may gain or lose marks in some subjective way.

Administration of the tests.

1. Set and administer test in your usual way. Pupil's answers should be on loose paper with name, class and subject clearly indicated.

2. At conclusion of each test (before answer papers are handed in) distribute an additional slip of paper to each child. Instruct pupils to write their name, class and date at the top of this slip. This is the pupil's assessment slip.
3. Instruct the pupils as follows: "I want you to imagine that you are going to mark your own test paper. Write down on your extra slip the mark, out of 20, that you think your test paper is worth". Please evade any questions from the pupils, but ensure that each pupil does write an assessment mark on his or her slip. Discourage any discussion between pupils.

Marking and Recording (See example sheet)

Test 1

1. After Test 1 has been marked, arrange the answer papers in order from top to bottom scoring pupil. Pupils with the same score should be ordered at random. Then commencing with the highest scoring pupil and proceeding down the class:

2. Record the pupils names and sex on the record forms.

3. Record the Actual mark scored for the test in the appropriate space on the form.

4. Record the Pupil's expected mark for the test in the appropriate space.

5. Determine what experimental treatment is appropriate for the test by reading the code mark for that pupil. Enter the resultant mark in the space marked R.

6. Enter this mark (R) on that pupils assessment slip, and initial the slip.

7. Retain all test answer papers and return the assessment slips to the pupils at the appropriate time before the next test.

Tests 2 - 6

Follow the procedures as outline in 3 - 7 above.

Test 7

This is the conclusion of the experiment and no further treatments are necessary. The only recording required is of the pupils actual and expected marks. The marks returned to the pupils should be their actual scores on Test 7.

General points.

1. Please retain all test answer papers.

2. Please give no indication to the children that they are taking part in an experiment.
Special conditions (See example sheet)

1. Absence.
   Enter 'Ab' in the appropriate spaces.
   (See 'R Heap' on the example sheet).

2. No Treatment.
   There will be rare occasions when pupils will fall into a treatment category where it would be unlikely or impossible to achieve the 'treatment score' based on their expectations. Thus (1) a pupil in the low treatment group who expects to receive 2 marks or less should be returned his or her actual score and N.T. written in the appropriate space. Similarly (2) a pupil expecting 18 marks or more in a 'High treatment group'

(See B. Fudge, Test 2; Y Ware Tests, 4.5.6; and B. Franklin, Test 5.).
<table>
<thead>
<tr>
<th>Name</th>
<th>Initial</th>
<th>Treatment Code I</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Treatment Code II</th>
<th>Test 4</th>
<th>Test 5</th>
<th>Test 6</th>
<th>Test 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. B. Franklin</td>
<td>F</td>
<td>-</td>
<td>18</td>
<td>18</td>
<td>16</td>
<td>+</td>
<td>18</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>2. R. Cherry</td>
<td>M</td>
<td>✓</td>
<td>17</td>
<td>16</td>
<td>17</td>
<td>=</td>
<td>17</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>3. R. Heap</td>
<td>M</td>
<td>✓</td>
<td>17</td>
<td>19</td>
<td>17</td>
<td>=</td>
<td>17</td>
<td>11</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>4. M. Dixon</td>
<td>F</td>
<td>-</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>+</td>
<td>16</td>
<td>14</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>5. N. Thomas</td>
<td>M</td>
<td>=</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>✓</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>6. R. Frances</td>
<td>F</td>
<td>+</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>=</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>7. B. Fudge</td>
<td>M</td>
<td>=</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>+</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>8. Y. Ware</td>
<td>+</td>
<td>✓</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>✓</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Key:**
A. Actual Mark
E. Expected Mark
R. Mark Returned

**Codes:**
- + Add 2 to Expected Mark
- - Subtract 2 from Expected Mark
= Return Expected Mark
✓ Return Actual Mark
TEACHERS’ MARKS AND PUPILS’ EXPECTATIONS: THE SHORT-TERM EFFECTS OF DISCREPANCIES UPON CLASSROOM PERFORMANCE IN SECONDARY SCHOOLS

BY

ANTHONY J. PICKUP AND W. S. ANTHONY

Off-printed from:

THE BRITISH JOURNAL OF EDUCATIONAL PSYCHOLOGY

NOVEMBER 1968

VOL. 38, PART 3

Price One Pound
TEACHERS' MARKS AND PUPILS' EXPECTATIONS: THE SHORT-TERM EFFECTS OF DISCREPANCIES UPON CLASSROOM PERFORMANCE IN SECONDARY SCHOOLS
TEACHERS' MARKS AND PUPILS' EXPECTATIONS: 
THE SHORT-TERM EFFECTS OF DISCREPANCIES UPON 
CLASSROOM PERFORMANCE IN SECONDARY SCHOOLS 

BY ANTHONY J. PICKUP AND W. S. ANTHONY 
(School of Education, University of Leicester)

SUMMARY. The research investigated (a) pupils' expected marks for a piece of work and their relationship to the actual mark for that work and (b) the effects of discrepancies between expected marks and teacher's marks. In classroom tests of normal classwork, it was found that actual marks tended to be less than expected marks, especially when actual mark was low. Discrepancies between pupils' expectations and teacher's mark were found to have no single motivational effect. However, there was a relationship between effectiveness of discrepancies and actual mark; the treatment of giving less marks than a pupil expected showed a slight tendency to be more effective than the opposite treatment when actual (objective) mark was high, but less effective when it was low. Thus it appears that the most usual treatment differed from the most effective treatment.

I.—INTRODUCTION.

When a teacher asks his pupils to perform tasks for which they will be returned marks or grades he is initiating an activity which is one of the most extensively applied methods in use in education. The situation which arises when a mark or grade is returned to a pupil is at first sight a simple application of an essential "feedback" process; a method by which the teacher can assess the success of his teaching and by which a learner can assess the adequacy of his learning. In the typical educational context however, the process of returning marks to pupils on completion of a task is not merely informational. The situation contains a number of elements which may affect the subsequent motivational state of the pupil. The simplest motivational element is the avoidance of the lack of interest which would result from no knowledge of results, but other elements are also present. A good mark may be considered a reward and a poor mark a punishment. Marks may be accompanied by statements or comments so that together or singly they constitute a situation of praise or of blame. The motivational state of the learner may be influenced by his perception of the likelihood of success or failure and the perceived status of the teacher may influence the reception of marks, grades or comments.

The now notorious unreliability of marks which may be awarded (Vernon 1962, Britton et al. 1966) may be further complicated by the teacher deliberately distorting marks for motivational purposes (Vernon 1962, p. 33). The practice of deliberately distorting marks introduces a dangerous assumption. The teacher who does this presumably makes a number of assumptions about the effects which a mark will have upon a pupil's later performance. Is a low mark [whatever that may mean] necessarily discouraging?

It is reasonable to assume that when a pupil has completed a task for which he is to receive a mark or grade, the pupil forms an expectation of some kind regarding this mark or grade. This expectation may be of a crude nature; a 'good' mark or a 'poor' mark. It may be further refined in relation to the pupil's past performance; a "better mark than usual" or a "worse mark than usual." If the teacher asks his pupil to state what he (the pupil) considers
his work is worth, this evaluation may be the closest approximation to the pupils' true expectation that can be found. Any vague 'hopes' which may be entertained by the pupil will be modified by some realism, because the recipient of his evaluation is the teacher who will mark his work. This evaluation is therefore described throughout the present report as the pupil's expectation.

The aims of the research here described were to examine, in typical classroom settings, the following:

(a) The relationship between pupil's expectations and the teacher's mark.
(b) The effects upon subsequent performance of a discrepancy between the pupil's expectation and the mark awarded by the teacher.

II.—Previous related work.

A search of the literature concerned with educational incentives revealed no previous work directly related to this topic. While a number of analogies may be drawn with research in the area of "praise or blame," these do not present precise guidelines for the formation of hypotheses. Kennedy and Willcutt (1964) in their review of literature in this area, have shown that the experimental history of praise and blame studies has failed to produce clear-cut conclusions, though recent evidence appears to support the belief that praise is a superior incentive to blame. "Level of aspiration" studies offer further analogies, limited by the fact that goal-setting is a pre-performance factor while the present research is concerned with the post-performance factor of expected mark. Child and Whiting (1949) illustrate a number of widely accepted generalisations in the area of level of aspiration. Murstein (1965), studying the relationship of grade expectations to grades actually received, found that students did not change their expectations of grades as a result of receiving actual grades. High ability students expected high grades and received them. Low ability students similarly expected high grades and did not modify their optimism as a result of experience.

The authors were impressed with the method of the Page (1958) research, concerned with the effectiveness of teachers' comments. This research has been popular with reviewers (Campbell and Stanley 1963, Charters and Gage 1963) because of its avoidance of many pitfalls in a 'generalisable' design. The experimental methods used in the research described below are evolved from the Page (1958) research.

III.—Plan of investigation.

Preliminary Work.

A six-class study, described in full elsewhere (Pickup 1967, ch. 6), was carried out to examine the relationship between teacher's marks and pupil's expectations which would occur in schools under non-experimental conditions. Measures were taken of pupils' expected marks following classroom tests and of teachers' marks for the tests. Precautions taken to avoid the 'artificiality' of experiment included the use of the usual classroom teacher as a sub-experimenter. The main findings of this study were as follows:

(1) Out of a final sample of 123 pupils three had estimated their mark correctly, relative to the teacher's mark.

(2) Of those pupils who were not accurate in their estimate 68.3 per cent over-estimated their mark and 31.7 per cent under-estimated their mark on the first occasion of being asked to evaluate their work. (This is a highly significant difference, P < 0.001.)
(3) Girls who had over-estimated their mark on the first occasion modified this optimism significantly more than over-estimating boys when both were asked to evaluate their work on a second occasion.

(4) Low pre-test scorers over-estimated more often than high pre-test scorers. (This was a highly significant difference, $P < .001$.)

Result (4) listed above, the correlation between pre-test score and frequency of over-estimation, arose in the following way: The data were examined for indications of a possible effect of discrepancy between expected and teacher's mark upon subsequent test performance. It was recognised that if there were a correlation between pre-test score and discrepancy, then the well-known phenomenon of statistical regression (Campbell and Stanley 1963, p. 181) would produce the false appearance of an effect of discrepancy upon subsequent performance. Therefore result (4) was looked for, and the actual finding of the result meant that this particular study would not yield a true picture of the possible effect of discrepancy. In any case, this was a correlational study and hence less satisfactory than an experimental study in elucidating cause-and-effect relationships. It was therefore decided to carry out an experimental study in which the discrepancy between expectation and test-score would be manipulated as an independent variable, while retaining (as far as possible) the real-life aspects of the situation.

A two-class pilot study was carried out to test a possible experimental design. This study (Pickup 1967, ch. 7) yielded useful information which led to the improved design described below.

**The Main Study.**

The major purpose of the main study was to examine by experimental means the effects upon subsequent classroom performance of a discrepancy between a teacher's mark and a pupil's expectation. A secondary purpose was to check the findings of the preliminary work described above.

Six non-selective single-sex secondary schools provided a final sample of 253 pupils from ten classes. The school subjects used in the experiment were English and Mathematics.

**IV.—The measures used.**

*Performance Tests.*

Participating teachers were requested to set two tests of their normal classwork. The material to be tested was to be as independent of previous learning and experience as possible. The teachers were asked to construct their tests as follows:

(a) A section of 20 'objective' questions which could be scored easily on a right/wrong basis.

(b) A 'woolly' question containing a subjective element. The pupil should not be able to assess easily whether marks are gained or lost on this question. (This question was never marked because its sole purpose was to disguise the fact that marks had been tampered with.)

(c) The tests were to be so constructed as to give every pupil an opportunity of gaining some marks, but extreme difficulty in obtaining full marks. These tests were to be marked by the experimenter, following a marking scheme devised by the participating teacher. Teachers were informed that the marks returned after the first test may differ from the actual performance of the pupil.
Pupils' Expected Marks.

Pupils were asked to enter upon an additional slip of paper the mark (out of 20) that they believed their test paper to be worth. These assessment slips were to be given to the experimenter together with the test scripts.

Additional Information.

The Junior Eysenck Personality Inventory (Eysenck 1965a, b) was administered to the pupils during a testing session separate from the experiment. A measure of the pupils' academic motivation was obtained during the same session by means of a modified 19-item test described by Buxton (1965). A measure of the participating teachers' leadership attitude was taken by means of a modified version of the Assumed Similarity of Opposites test described by Fiedler (1958).

V.—PROCEDURE.

Participating teachers were supplied with an instructional leaflet concerning their part in the experiment. After the administration of the pre-test and collection of the pupils' assessment slips, the experimenter collected this material and marked the pre-test scripts following the teacher's marking scheme; these marks are referred to as 'actual' marks. All pupils who had indicated that they had been absent for some portion of the work tested in the pre-test were assigned to a 'No Treatment' group. Similarly, all pupils who had recorded an expected mark of 0-3 or 17-20 were assigned to the No Treatment group. (However, in one class of 26, 10 pupils expected a mark of 17-20, and in this case only those pupils who would have fallen into the High Treatment group were excluded, which resulted in a loss of three pupils).

Pupils remaining after these exclusions were randomly assigned to one of three treatment groups (High, Low and Matching treatments) within levels of 'actual' score rank. Thus, within any one class-list, actual scores were converted into rank order; the three top ranks (1, 2, 3) were grouped into a triplet, the next three ranks (4, 5, 6) were grouped into a next triplet, and so on. Ties had previously been ranked by the toss of a coin. Two numbers on an ordinary six-sided die were selected to represent each of three treatments, for the purpose of random assignment: within each triplet, the first member was assigned to a treatment by a throw of the die; the second member was similarly assigned to a second treatment; and the third member was assigned to the remaining treatment. This procedure was repeated for each triplet.

The experimenter administered experimental treatments by writing, on a mark list, a mark which was 3 marks above ('High' treatment), 3 marks below ('Low' treatment), or exactly matching ('Matching' treatment) the pupil's expected mark, according to the treatment group in which that pupil fell. This mark list was returned, with the pupils' assessment slips, to the participating teacher. The teacher entered the mark from this list on to each pupil's assessment slip and this was returned to the pupil (i) immediately prior to the work which was to be tested on the post-test or (ii) immediately prior to the post-test where (i) was not possible. (An excuse was made to the pupils that the actual test scripts were not available at that moment; in fact neither pre-test nor post-test scripts were ever returned to the pupils.)

Post-test scripts, marking scheme and assessment slips were collected by the experimenter and a post-test mark list, based upon the teacher's marking scheme, was returned to the teacher with the assessment slips. During this
visit to the school, the experimenter administered the personality and academic motivation tests to the pupils while the participating teacher completed the ASO test in another room. All pupils were eventually informed that an experiment had taken place and that the scripts were required for records.

The percentile rank of each raw score in a class test was calculated and converted into a T score (Garrett 1958, pp. 314-318, Table G). Thus each set of test scores has a normal distribution with a mean of 50 and a standard deviation of 10.

VI.—Results.

A.—Pupils' expected mark and actual mark.

A1. When the pre-test data were examined it was found that 72.6 per cent of the pupils over-estimated their mark relative to the actual mark. A considerable number of pupils (almost 10 per cent) were accurate in their estimate. Of those who were not accurate in their estimate, 81.3 per cent over-estimated their mark. This tendency to over-estimate is highly significant (P < .001) and confirms the finding of the preliminary work.

A2. The data were examined for significant direction of change of expectation following experimental treatment. It was found that girls who been exposed to Low treatment following the pre-test showed a highly significant tendency (P < .001) to reduce their expectations for the post-test. No other significant directions of change were found.

A3. Pre-test score was correlated with the difference between actual mark and expected mark (r = .33, P < .001). That is, low test scorers were especially liable to over-estimate their marks, as had also been found in the preliminary work.

B.—The effects of the experimental treatment.

B1. The effects of each treatment upon gain scores were examined by means of the analysis of variance by ranks. Each intact triplet in the experiment yielded ranks for a Friedman two-way analysis of variance (Siegel 1954, pp. 166-172). The ranked data from each triplet were cast into various categories to examine treatment effects. There were no significant effects here. Other analyses, carried out to examine possible subject differences, also proved non-significant.

B2. Analyses concerned with the exploration of treatment effects in relation to personality factors, academic motivation and teachers leadership attitude, proved non-significant except for some minor sex-differences.

B3. Examination of the data on High and Low treatment and pre-test position in class showed a correlation between pre-test score and differential effect of treatment. It is to be noted that in this experimental design, treatments are applied with equal frequency within each level of pre-test score, and therefore there is no misleading effect of statistical regression as there was in the earlier correlational study. This correlation was at first tested by relatively crude techniques which failed to yield significant results, and then by the product-moment method, as follows. The variables studied were matched-pairs difference scores and mean pre-test scores. Each randomised triplet in the experiment was examined for losses of individuals; any triplet which had lost a high or low treatment pupil was excluded from analysis. (i) The standardised scores for the high treatment pupil and the low treatment pupil yielded a mean pre-test score for that pair. (ii) The gain score for the high treatment
pupil was subtracted from the gain score for the low treatment pupil to yield a difference score for that pair. A product-moment correlation between the variables in (i) and (ii) yielded $r = 0.275$ $(N=74$ pairs). The significance of this obtained $r$ was tested (Walker and Lev 1953, p. 251) yielding $t = 2.43$, df. = 72. This correlation is significant beyond the .02 level (two-tailed test).

This significant correlation indicated the need to examine all differences between treatments in the same way. No significant differences were found other than between High and Low treatment.

A crude indication of treatment effects related to pre-test score is to be found in Figure 1. In this graph three performance groups (pre-test) 'failing' pupils (45 marks and less) 'average' pupils (between 46-54 marks) and 'successful' pupils (55 marks and more) are studied to show average gain scores following experimental treatment. The general downward slope is a reflection of statistical regression, but the significant aspect of the figure is the difference between experimental treatments at the various levels of pre-test score.

![Figure 1](image-url)

**Figure 1**

*The Effects of the Three Experimental Treatments at Different Levels of Pre-Test Score. The Gain Score is the Post-test $T$ Score Minus the Pre-test $T$ Score.*
VII.—Conclusions and Discussion.

This research aimed to investigate (a) the relationship between pupils' expected marks and actual marks and (b) the effects, upon subsequent performance, of discrepancies between expected marks and marks received by the pupil from the teacher. The method of the research was designed for generalizability to the normal school setting; the only divergence from normal classroom procedure was in the collection of pupils' expected marks. It was intended, and it is believed that pupils were unaware that observation or experiment were taking place.

The results naturally fall into two categories corresponding to the two aims of investigation.

(a) Concerning the relationship between expected and actual marks, it was found that (1) a large majority of pupils over-estimated their mark on the first test; (2) girls who expected more marks than they received showed a tendency to reduce their expectation on the second test; (3) low test scorers were especially liable to over-estimate their mark.

(b) Concerning the effects upon performance of discrepancies between expected and received mark, it was found that (1) there was no significant overall effect of discrepancies; (2) there was no significant effect of discrepancies when school subject, sex of pupil, personality of pupil or teacher's leadership attitudes was taken into account, except for some minor sex-differences; (3) effect of discrepancy was significantly related to pre-test score: giving a pupil more (rather than less) marks than he expected was found to be the more effective treatment of the 'poorer' pupils but the less effective treatment of the 'better' pupils.

Results (a) (3) and (b) (3) should be considered together. Result (a) (3) and the preliminary work, indicated that low test scorers especially score less than they expect. Result (b) (3) suggests that it would be a more effective treatment to give low test scorers more than they expect; high scorers seem to perform better if they have been given less than they expect. So far as the low scorers at least are concerned, it appears that the better treatment is in the opposite direction to what tends to occur. Any beneficial effect of a novel treatment may, of course, be due only to its novelty.

Implications for further research.

From the present study it seems that if there is a single motivational effect of a discrepancy between a pupil's expected mark and the mark received from the teacher, this effect is so small as to be insignificant in the normal classroom setting. Such an effect might, however, appear with the use of more sophisticated performance tests and laboratory-type conditions. Other experiments employing a method similar to that employed in the main study could investigate the effects of repeated discrepancies over a longer period of time. Some participating teachers suggested that the procedure of asking a pupil for a written assessment of his own work was in itself a motivating influence; this is worth investigating.

It seems that the effectiveness of one treatment relative to its opposite, as an incentive, reverses as the opposing treatments are applied at different levels of test-score in the class. This 'reversal effect' could conceal a treatment effect if positive effects of a treatment at top ability levels are 'cancelled-out' at the bottom levels. In no study of educational incentive which has been found has the possibility of this reversal effect been considered. Further investigation of the effect would be desirable.
ACKNOWLEDGEMENTS.—Thanks are due to the Chief Education Officers of the London Boroughs of Croydon and Bromley and to the head teachers and participating staff of the ten schools in which the research took place. Thanks also are due to Dr. Sybil B. G. Eysenck, who kindly supplied the Personality Inventories used.

VIII.—References.


(Manuscript received 28th March, 1968)
OTHER PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Psychology. The British Journal of Medical Psychology.
The British Journal of Social and Clinical Psychology.

The British Journal of Psychology and The British Journal of Medical Psychology appear in quarterly parts in volumes of about 350 pages, subscription price of each, payable in advance, 60s. per volume (post free). Subscriptions may be sent to any Bookseller or to the Cambridge University Press, Bentley House, 200, Euston Road, London, N.W.1. The subscription price in the U.S.A. is $10.00. Enquiries should be addressed to the Cambridge University Press, American Branch, 32, East 57th Street, New York 22.

The British Journal of Mathematical and Statistical Psychology is issued twice a year, in May and November, in volumes of about 184 pages. The subscription rate payable in advance is 30s. 6d. per volume (post free). The subscription price in the U.S.A. is $4.50 (post free). Subscriptions should be sent to the Society's Agents, Taylor and Francis, Ltd., Red Lion Court, Fleet Street, E.C.4, or to any Bookseller.

The British Journal of Social and Clinical Psychology is issued three times a year in February, June and November, in volumes of approximately 240 pages. The subscription price of each, payable in advance, is 50s. (post free). Subscriptions may be sent to any Bookseller, or to the Cambridge University Press, 200, Euston Road, London, N.W.1. The subscription price in the U.S.A. is $8.50. Enquiries should be sent to the Cambridge University Press, American Branch, 32, East 57th Street, New York 22. Separate parts will be sold at 20s. : $3.50 in the United States.

A series of Monograph Supplements is issued in connection with the British Journal of Psychology, but is not included in the subscription for that journal. Orders may be sent to any Bookseller or to the Cambridge University Press, 200, Euston Road, London, N.W.1, from whom details of these publications may be obtained.

MANUSCRIPTS

For the British Journal of Psychology should be sent to Professor W. SLUCKIN, Department of Psychology, The University, Leicester.

For the British Journal of Medical Psychology, should be sent to Dr. THOMAS FREEMAN, Lansdowne Clinic, 4, Royal Crescent, Kelvingrove, Glasgow, C.3.

For the British Journal of Educational Psychology, should be sent to Professor J. D. NISEET, King's College, Aberdeen.

For the British Journal of Mathematical and Statistical Psychology, should be sent to Dr. R. J. AUDLEY, Department of Psychology, University College, Gower Street, London, W.C.1.

For the British Journal of Social and Clinical Psychology, should be sent to Mr. MICHAEL ARGYLE (Social Psychology Editor), Institute of Experimental Psychology, 1, South Parks Road, Oxford;

or to:

Dr. RALPH HETHERINGTON (Department of Studies in Psychological Medicine), 6, Abercromby Square, Liverpool, 7.

For the Monograph Supplements, to Professor IAN M. L. HUNTER, Department of Psychology, University of Keele, Keele, Staffs, whose decision regarding the publication of all Monograph Supplements will be final. No MS. can be considered for publication in any year unless it is in the hands of the Editor by 1st September of the preceding year.

OFF-PRINTS

Contributors receive 50 free copies of their article and may buy more in units of 50. Members of the British Psychological Society receive all the Society's Journals at reduced rates.

Information on membership and further activities may be obtained from the Secretary, Tavistock House South, Tavistock Square, London, W.C.1.

PUBLISHED BY THE ASSOCIATION OF TEACHERS IN COLLEGES AND DEPARTMENTS OF EDUCATION

Education for Teaching is published three times a year at mid-term in November, February and May, and distributed free to members of the Association. To non-members it costs 5s. per issue, post free (subscription 14s. per annum).

Members of the A.T.C.D.E. receive the British Journal of Educational Psychology at a reduced rate.

Information on membership and further activities may be obtained from the Honorary Secretary, 151, Gower Street, London, W.C.1.
CONTENTS

H. B. Gibson:
Parental Attitudes and their Relation to Boys' Behaviour 233

K. Lovell and A. Gorton:
Some Differences between Backward and Normal Readers of Average Intelligence 240

D. Joynt and B. Cambourne:
Psycholinguistic Development and the Control of Behaviour 249

E. Stones and J. R. Heslop:
The Formation and Extension of Class Concepts in Primary School Children 261

H. J. Butcher and H. B. Pont:
Opinions about Careers among Scottish Secondary School Children of High Ability 272

J. H. Clarke:
The Image of the Teacher 280

P. K. Poppleton:
Puberty, Family Size and the Educational Progress of Girls 286

B. P. Frost:
Anxiety and Educational Achievement 293

A. J. Pickup and W. S. Anthony:
Teachers' Marks and Pupils' Expectations 302

Research Notes:
S. Lemeshow:
A Problem Check List for Teachers 310

C. K. Mackay and T. M. Kilkenny:
The Stability of Induced Conservation 313

C. K. Mackay and M. B. Cameron:
Cognitive Bias in Scottish First-year Science and Arts Undergraduates 315

F. Fee:
An Alternative to Ward's Factor Analysis of Wallach and Kogan's 'Creativity' Correlations (Comment by J. Ward) 319

Theses Abstracts:
R. Hume:
The Effect of School and Curriculum on Boys' Attitudes towards Careers in Engineering 322

R. Amos:
Teachers' Opinions on the Place of Scientific Method in Biology Courses 323

C. L. Baird:
The Role of the Teacher of Six- and Seven-year-old Children 323

B. Staveley:
The Abilities and Interests of Craft and Technician Students of Mechanical Engineering 324

Book Reviews 325

Publications Received 336

Index to Volume 38 338

Notes to Contributors 341
REFERENCES


MURSTEIN, B.I. (1965) "The relationship of grade expectations and grades believed to be deserved to actual grades received". J.Exp. Educ. 33, 357-362.


