DOCUMENT RESUME

ED 083 345 UD 013 866

AUTHOR	Dusek, Jerome B.
TITLE	Teacher and Experimenter Bias Effects on Children's Learning and Performance.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	28 Aug 73
GRANT	OEG-2-71-0516
NOTE	39p.; Paper presented at the American Sociological Association annual meeting, New York, N.Y., August 28, 1973
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Academic Performance; Achievement Tests; *Bias; Childhood Attitudes; Classroom Environment; Educational Experiments; *Elementary Grades; Expectation; Interaction Process Analysis; *Learning; Longitudinal Studies; Predictor Variables; *Teacher Attitudes
IDENTIFIERS	SAT; Stanford Achievement Test

ABSTRACT

Three experiments were conducted to examine the effects of adult expectations on children's learning and performance; one in-classroom study and two experimental studies were made in order to investigate developmental trends in susceptibility to expectancy effects and the relationship of induced vs. self-generated expectancies vis-a-vis children's learning and performance. The major experiment was a 1-1/2 year longitudinal study of teacher bias and expectancy effects on the Stanford Achievement Test (SAT) performance of children in two grade 2 and grade 4 classrooms. The major findings were that: telling teachers students will do well did not alter children's SAT performance; teacher ranking was significanlty related to SAT performance from each of the five testing periods; and, there were no interactions with grade level. These findings were interpreted as indicating that teachers are good predictors of children's academic potential and do not "bias" children's education. The finding of the grade level X experimenter sex X sex of Ss X time study of experimenter bias was the significant triple interaction involving grade level, bias condition, and sex of Ss; this interaction reflected a general trend of older Ss to be more influenced by biasing effects of experimenters than younger Ss. The second experimental study revealed essentially the same effects for experimenters in who bias was induced and those who predicted performance themselves (self-generated bias). (Author/RJ)

ERIC

FILMED FROM BEST AVAILABLE COPY

ED 083345

U S DEPARTMENT OF HEALTH.
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Teacher and Experimenter Bias Effects on Children's Learning and Performance [1,2]

Jerome B. Dusek

Syracuse University

Paper presented at the 68th Annual Convention of the American Sociological Association, New York, New York, August 28, 1973.

UD 013866

Teacher and Experimenter Bias Effects on Children's Learning and Performance

Jerome B. Dusek

Syracuse University

Abstract

Three experiments were conducted in order to examine the effects of adult expectations on children's learning and performance. One in-classroom study and two experimental studies were conducted in order to investigate developmental trends in susceptibility to expectancy effects and the relationship of induced vs. self-generated expectancies vis-a-vis children's learning and performance. The major experiment was a 1½ year longitudinal study of teacher-bias and teacher-expectancy effects on the Stanford Achievement Test (SAT) performance of children in two second- and two fourth-grade classrooms. The major findings were: (a) telling teachers students will perform well did not alter children's SAT performance; (b) teacher ranking was significantly related to SAT performance from each of the five testing periods; (c) there were no interactions with grade level. These findings were interpreted as indicating that teachers are good predictors of children's academic potential but do not "bias" children's education. The major finding of the 3 (Grade Level) x 2 (Sex of Experimenter) x 2 (Sex of Subject) x 7 (Minutes) study of experimenter-bias in a simple motor performance task (marble dropping) was the significant triple interaction involving Grade Level (1st, 3rd, 5th), Bias Condition, and Sex of Subject. This interaction reflected a general trend for older subjects to be more influenced by biasing effects of experimenters than younger subjects. The second experimental study, a 2 (Induction Condition) x 2 (Bias Condition) x 2 (Sex of Subject) x 7 (Minutes) design, revealed essentially the same effects for experimenters

in whom the bias was induced and those who predicted performance themselves (self-generated bias).

The central problem under investigation was the effect of adult/teacher expectations on children's learning and performance. Three studies were conducted to provide information relevant to the following three questions: a) Are teacher-bias or teacher-expectancy effects observable in measures of academic performance? b) Are these effects observable only when induced in the teacher or experimenter by the principal investigator as opposed to being self-generated by the adult? c) Are there developmental trends in susceptibility to adult expectancy effects?

Research bearing on these issues falls into three categories. First, there is a body of research dealing with experimenter bias effects in psychological research. This research has been thoroughly reviewed by Rosenthal (1966, 1968, 1969a,b), Friedman (1967) and Barber and Silver (1968a, 1968b). The literature in this area is a clear demonstration that under certain conditions experimenters may intentionally or unintentionally bias the performance of adults (Rosenthal, 1966; Barber & Silver, 1968a, 1968b) or children (Dusek, 1971; 1972) in psychological experiments. Second, there are several studies in which expectancy effects and self-fulfilling prophecies have been investigated in tutoring situations involving student teachers (e.g., Beez, 1968; Rubovits & Maehr, 1972). Third, there are a number of studies in which teacher expectancy effects in elementary school classrooms, or other classroom situations, have been investigated (e.g., Rosenthal & Jacobson, 1968; Claiborn, 1969).

In the remainder of this paper the term "teacher-bias or experimenter-bias effects" will refer to significant effects due to teacher/experimenter differential expectations for children's performance, but only in the case involving induction of expectancies by a principal investigator. That is, bias effects will be due to a manipulation, or attempted manipulation, of expectancies by an investigator. ...ch effects are analagous to the effects reported by Rosenthal (1966) and

Rosenthal and Jacobson (1968) and are bias in the sense that the adult has differential expectations regarding the performance of children who are equivalent on some objective measure. The term "expectancy effects" will refer to significant effects due to the adults' own, self-generated expectations regarding children's performance. In this case, it is the adults' own expectancy, formed however adults form it, which is related to children's performance. This distinction will prove critical in interpreting the findings reported below.

The first study of teacher bias effects was conducted by Rosenthal and Jacobson (1968) in an elementary school serving primarily a lower social class neighborhood. At the beginning of the school year all the children in grades 1-6 were given an IQ test, Flanagan's (1960) Test of General Ability (TOGA), disguised as a test to predict "academic blooming". The test was given again at the middle and end of the school year. Within each of the 18 classrooms approximately 20% of the children were randomly chosen to form an experimental group. The names of these students were given to their teachers and it was explained that these children had scored on the test in such a manner as to predict that they would show large gains in intellectual ability during the school year. Across all classrooms the year-end test scores showed an approximately 4 point advantage for the children in the experimental group. However, at the first- and second-grade levels the children in the experimental group showed gains of as much as 15 IQ points more than the children in the control group. In terms of school performance, the children in the experimental group showed a significantly better gain than the children in the control group only for reading, one of the 11 school grades considered. On the basis of these data Rosenthal and Jacobson (1968) concluded that the ·ildren in the experimental group gained more than children in the control p during the course of the academic year because the teachers expected a higher level of performance from them.

R. L. Thorndike (1968, 1969) has criticized the Rosenthal and Jacobson research on several grounds, including faulty pre- and post-test data and the suggestion that students may not have attempted a large number of items, thus lowering their IQ scores and, essentially, making the test a poor measure. Jensen (1969) has attacked the Rosenthal and Jacobson research on three grounds: a) the same IQ test was used for the pre- and post-tests; b) the teachers administered the tests; c) the child was the unit of analysis instead of the classroom. In addition, Claiborn (1969) has argued that many of the findings are unconvincing since they did not reach standard levels of significance and were not predicted prior to the investigation. Rosenthal (1968; 1969b; 1973) has convincingly replied to many of these criticisms. However, other criticisms of the Rosenthal and Jacobson research remain (e.g., see Elashoff & Snow, 1970, 1971; Rosenthal & Rubin, 1971). As Snow (1969) has argued, "Rosenthal and Jacobson will have made an important contribution if their work prompts others to do sound research in this area. But their study has not come close to providing adequate demonstration of the phenomenon or understanding of its process." At the present time, no other conclusion regarding this research seems reasonable or possible. The Rosenthal and Jacobson research has stimulated a number of studies exploring various aspects of teacher bias, or self-fulfilling prophecy, effects, however. It is these which shall be reviewed next.

Claiborn (1969) attempted to replicate the Rosenthal and Jacobson research. Not only were some of the teachers led to believe certain students would show much progress intellectually during the remainder of the year, but some of the classrooms were observed in order to obtain data concerning the student-teacher interactions. The results indicated no differential gain in IQ between the experimental and control children. Furthermore, there was no indication that teachers behaved differently toward the control and experimental children. Since the biasing statements were introduced well into the school year (Spring semester) and since the length of the study was only 2 months, the results are difficult to interpret. Perhaps the teachers had their own well-formed opinions of the students' potential and the opinion of an "outsider" was just not seen as valid. Perhaps, too, the two month interval between biasing and post-testing was not long enough for the effect of teacher expectancy to become critical to the students' performance.

Several other studies have attempted to replicate the findings of Rosenthal and Jacobson with varying degrees of success. Evans and Rosenthal (1969) found that for Kindergarten through fifth grades the boys in the experimental group gained more IQ points in the reasoning subtest than the boys in the control group, with the reverse holding for the girls. There were no effects for Verbal IQ or Total IQ scores. Anderson and Rosenthal (1968) report a failure to replicate with familial retarded boys. Meichenbaum, Bowers, and Ross (1969), using female adolescent offenders, reported that the "bloomers" showed more improvement on objective but not on subjective tests than did the control group. This study is of particular interest since it focused on academic performance rather than IQ. Furthermore the classroom observations revealed that the "bloomers" significantly improved in terms of appropriate behavior more than did the control group.

Experiment I

A Longitudinal Study of Teacher-bias and Teacher-expectancy Effects on Elementary School Children's Achievement Test Performance

The purpose of this experiment was to investigate teacher-bias and teacher-expectancy effects on elementary school children's achievement test performance. Teacher-bias is defined as above, that is, an expectancy for performance as induced by the principal investigator. Analagously, teacher expectancy is defined as the teacher's own self-generated (stated) expectations regarding children's performance. In this experiment, as will be noted below, teacher-bias was manipulated by statements from the principal investigator and teacher-expectancies were measured by teachers' rankings regarding year-end academic performance levels.

Subjects

The subjects were 32 second-graders ($\overline{CA}$ = 8.60 years), 13 boys and 19 girls, and 32 fourth-graders ($\overline{CA}$ = 10.73 years), 15 boys and 17 girls, attending a school serving primarily a lower-class population. There were 16 subjects in each of two classroom in each grade level.

Procedure

During the first week of the 1971-1972 academic year several subtests from the SAT battery were administered by the principal investigator to each of the classrooms involved in the study. The subtests administered included: Word Reading, Paragraph Meaning, Spelling, Arithmetic Computation, and Arithmetic Concepts (fourth-grade only). The Primary I and Partial Intermediate I batteries were used for the second- and fourth-grades, respectively. The SAT's were disguised as tests to measure potential gains in language and arithmetic skills.

The same subtests were administered at the middle and end of the 1971-1972 academic year. The SAT's were again administered at the beginning and middle of the 1972-1973 academic year, the children now being in the third- and fifth-grades. Subtests from the Primary II and Partial Intermediate I were now employed for the third- and fifth-graders respectively. It is important to keep in mind that the subjects were, at this time, in new grade levels with new teachers.

During the initial testing session each teacher was asked to rank the children in her classroom from 1-n based on her expectations regarding their year-end performance levels in language and arithmetic skills. In each classroom the children ranked 1-16 were randomly and equally divided into an experimental and a control group. One week after the initial testing each teacher was given the names of the children in the experimental group and was told that, on the basis of the tests, these children should show large gains in language and arithmetic skills during the academic year. It should be noted that no further mention of these children was made to any teacher throughout the remainder of the study, a year and a half.

Results

The dependent variables were total SAT raw scores for each testing session. Originally, the design was conceived as a three-way factorial arrangement, including experimental vs. control groups, grade level, and teacher ranking. However, due to subject attrition there were not an equal number of subjects in each cell.[3] Rather than solve the analysis problem by application of the unweighted means solution to the analysis of variance the multiple regression approach of Cohen (1968), Overall and Spiegel (1969), and Overall (1972) was employed.

The results of the multiple regression analyses are summarized in Table 1. The means associated with the main effects of the multiple regression analyses are presented in Table 2. As may be seen in Table 1 the bias manipulation (Experimental Condition) was not significantly related to SAT performance on any of the five testing occasions. Grade Level was significantly related to performance on SAT-2, SAT-3, and SAT-4. As may be seen in Table 2, the younger Ss scored higher than the older Ss on SAT-2 and SAT-3 with the reverse being the case for SAT-4.

Teacher ranking was strongly and consistently related to SAT performance on each testing occasion. In general, the higher the teacher's ranking the higher the child's SAT performance (see Table 2). The correlations between SAT performance and Teacher Ranking, presented in Table 3, reflect the strength of the relationships detected in the multiple regression analyses.

Conclusions

The findings are quite conclusive with respect to the importance of teacher-bias and teacher-expectancy effects on children's academic performance. Clearly, simply telling teachers certain students would be performing well at the end of the academic year was not sufficient to increase these students' SAT performance. It appears that the teachers biased neither the SAT performance nor the classroom learning of the children in the experimental or control groups. This appears to be the case for both short- and long-term effects due to teacher-bias. This finding does not replicate the findings of Rosenthal and Jacobson (1968). When considered in conjunction with other research (e.g., Claiborn, 1969; Anderson & Rosenthal, 1969; Evans & Rosenthal, 1969; Fleming & Anttonen, 1971; Jose & Cody, 1971) which has also failed to replicate the Rosenthal and Jacobson findings, however, it seems quite clear that teachers do not bias students' performance.

Table 1

Multiple Correlations and F-ratios for Each Condition at Each Testing Session

Source		SAT-1[a] (Oct., '71)	SAT-2[a] (Feb., '72)	SAT-3[a] (June, '72)	SAT-4[b] (Sept., '72)	SAT-5[b] (June, '73)
Experimental Condition	R^2 Full	.3536	.5472	.6825	.4229	.4598
	R^2 Reduced	.3536	.5465	.6817	.4213	.4537
	Difference	.0000	.0006	.0007	.0016	.0061
	F =	<1	<1	<1	<1	<1
Grade Level	R^2 Full	.3536	.5472	.6825	.4229	.4598
	R^2 Reduced	.3482	.4467	.3051	.3093	.4594
	Difference	.0055	.1005	.3774	.1136	.0004
	F =	<1	10.431*	55.852*	6.69*	<1
Teacher Ranking	R^2 Full	.3536	.5472	.6825	.4229	.4598
	R^2 Reduced	.0202	.1395	.4244	.0714	.0105
	Difference	.3334	.4076	.2580	.3516	.4493
	F =	24.243**	42.310**	38.188**	20.71**	28.28**

* $p < .01$
** $p < .001$
a df = 1/47
b df = 1/34

ERIC

Table 2

Mean Stanford Achievement Test Scores for Each Condition at Each Testing Session

Condition	Mean Test Score SAT-1	SAT-2	SAT-3	SAT-4	SAT-5
Condition					
Experimental	58.35	79.15	96.30	75.53	84.21
Control	56.63	76.51	96.71	69.58	81.11
Grade Level[a]					
Second (Third)	57.59	85.63	111.38	65.09	84.91
Fourth (Fifth)	57.38	70.37	79.95	81.63	79.56
Teacher ranking[b]					
Rankings 1-4	75.56	96.55	116.13	101.75	116.50
Rankings 5-8	60.13	81.48	100.82	67.13	86.50
Rankings 9-12	52.50	73.21	89.08	76.36	77.82
Rankings 13-16	41.75	60.06	80.00	49.73	60.09

[a] The grade level listed in the parentheses refers to SAT-4 and SAT-5.

[b] Teacher ranking was entered as a continuous variable in the multiple regression analyses. The data are grouped here simply for convenience.

ERIC

Table 3

Correlations Between Teacher Ranking and SAT Performance Across All Grade Levels and Conditions

SAT	r
1	-.59
2	-.67
3	-.55
4	-.55
5	-.67

Note. - $n = 51$ for SAT-1, SAT-2, and SAT-3 and $n = 38$ for SAT-4 and SAT-5. All *r*'s are statistically significant ($p < .001$).

Teacher Ranking was related to SAT performance on each testing occasion. Children ranked higher by the teacher had higher SAT scores than children ranked lower. This effect has been deemed a teacher expectancy effect since it reflects the teacher's own self-generated expectancy for the child's performance.

There is some evidence in the present study which supports the argument that this teacher-expectancy effect is not a teacher-bias effect in the Rosenthal and Jacobson (1968) sense. The first piece of evidence is the correlation between Teacher Ranking and SAT-1 performance. If this teacher-expectancy effect were due to teachers somehow biasing the test performance of the children it is unlikely that the magnitude of the correlation would have been as large. Second, teacher ranking was related to SAT performance 12 and 18 months after the ranking was made, the students now being advanced one grade level and under the tutelage of a new teacher. It is unlikely that this could be the case were the relationship based on a biasing influence by the teacher of the students' performance. Finally, the teachers reported that their rankings were based on criteria directly relevant to academic abilities, e.g., previous grades, readiness tests, and current classroom performance.

These effects due to teacher-expectancy appear to reflect the teacher's ability to accurately estimate the relative academic ability of the children in her classroom. The longitudinal data presented above appear to support this contention. Future research should focus on determining the exact bases used by teachers to form expectancies regarding students' abilities and the relationship of these bases to actual student performance as well as to teacher-student interaction in the classroom. Such research will not only clarify the nature of teacher-expectancies but also the role of the teacher in the child's cognitive and social development.

In order to gain a relatively complete understanding of children's reactions to adult expectations two studies aimed at examining experimenter-bias effects were also undertaken. Since this literature has been reviewed in several readily available sources (e.g., Rosenthal, 1966; Barber & Silver, 1968a, 1968b), and since the area is somewhat tangential to the major purpose of this symposium, the procedures and data will be presented in an abbreviated form. A more detailed report is available from the author.

Experiment II

A Developmental Study of Experimenter Bias Effects with Children as Subjects

Although E-bias effects have been shown in studies using children as Ss (e.g., Dusek, 1971, 1972), no information regarding developmental trends in susceptibility to E-bias effects is available. The major purpose of this experiment was to test for possible developmental trends.

Subjects

The subjects were 48 first ($\overline{CA}$ = 7 yrs. 4 mo., SD = 9 mo.), 48 third- ($\overline{CA}$ = 9 yrs. 5 mo., SD = 7 mo.), and 48 fifth-graders ($\overline{CA}$ = 11 yrs. 6 mo., SD = 8 mo.). Half the children in each grade level were males and half were females. The children attended a school serving primarily a lower-class neighborhood.

Experimenters

The experimenters were six male and six female college students ($\overline{CA}$ = 19 yrs. 11 mo., SD = 7 mo.) enrolled in the introductory psychology course at Syracuse University. Each E participated in both a group and an individual training session prior to testing the children (see below). During the experiment each E tested two boys and two girls from each grade level. Half the experimenters of each sex were randomly assigned to each bias condition.

Apparatus

The apparatus has been described in detail elsewhere (Stevenson & Fahel, 1961). Briefly, it consisted of a table with two bins and a transverse upright panel which served as a shield. The left bin contained approximately 1000 marbles. The table top above the right bin contained five randomly placed holes through which the marbles could be dropped. An Esterline Angus Event Recorder, shielded from S's view, was connected to microswitches below the holes and was used to obtain an automatic and permanent record of S's responses. The experiment was conducted in an area of the school free of distractions.

Procedure

Experimenter training. Each experimenter was randomly assigned to one of the two bias conditions. All experimenters (n=6) in the same bias condition attended the same group training session. The experimenters were shown the apparatus and the procedure was briefly outlined and demonstrated. The experimenters were then told they would be testing children in the public schools and the following biasing statement was made:

> We have used this task with this age children before and it has been found to be a sensitive measure of children's motivation. In fact, previous research shows that one of the findings we should expect to get is that boys (girls) will drop the marbles faster than girls (boys).

The procedures were then demonstrated again and each experimenter practiced the task. Each experimenter subsequently met with a graduate assistant to further practice the procedures.

Experimental task. The experimenter brought the subject to the testing room and read the instructions telling the subject to pick the marbles up one at a time and put them into the holes. As the subject picked up the first marble the experimenter started a stop watch and allowed the child to perform at the task for seven minutes.

During the first or baseline minute of the task the experimenter remained an attentive but nonresponsive observer of the subject's performance by glancing at the marbles and holes while avoiding looking at the subject. During the next six minutes, the experimental period, the experimenter used verbal reinforcers on a Fixed Interval 30-second schedule contingent on a marble drop. Six reinforcing statements were used: Good, Fine, That's good, That's fine, Very good, Very fine. Each subject received each statement twice in a predetermined random order. Each experimenter tested two boys and two girls at each grade level using this procedure.

Design

The experimental procedures required a 3 (Grade Level) x 2 (Sex of _E_) x 2 (Bias Conditions) x 2 (Sex of _S_) x 7 (Minutes) analysis of variance design with six subjects in the smallest cell.

Results

Dependent measures. There were two dependent variables of interest in the study: the base rate of response (the number of marbles dropped in the first minute of the task) and a series of difference scores computed separately for each subject by subtracting the number of marbles dropped in the first minute from the number of marbles dropped in each subsequent minute (Minutes 2-7). The correlation between the base-rate score and the average difference score was -.4133 (n = 144, $p < .01$) indicating that although the two variables are correlated only 17.1% of the variance in the difference scores is accounted for by the initial base rates.

Analysis of base-rate scores. The base-rate scores were subjected to a 3 (Grade Level) x 2 (Sex of Experimenter) x 2 (Bias Condition) x 2 (Sex of Subject) analysis of variance (see Table 4). The mean base-rate scores for each main effect are presented in Table 5.

As may be seen in Table 4, there were two significant effects. The significant Grade Level effect reflected a general increase in base rates with increasing grade levels. Newman-Keuls comparisons (Winer, 1962, p. 80) revealed that the means for each grade level were significantly different from each other (all $p < .01$). The Bias Condition x Sex of Subject interaction was also significant. The means are presented in Table 6. Individual comparisons (Winer, 1962, p. 207ff) revealed a

Table 4

Analysis of Variance of Base-Rate Scores

Source	df	MS	F	P
Grade Level (A)	2	604.000	23.16	<.001
Sex of *E* (B)	1	4.000	<1	
Bias Condition (C)	1	103.313	3.96	
Sex of *S* (D)	1	.438	<1	
A x B	2	14.656	<1	
A x C	2	56.188	2.16	
A x D	2	14.094	<1	
B x C	1	1.813	<1	
B x D	1	8.000	<1	
C x D	1	277.813	10.65	<.001
A x B x C	2	31.375	1.20	
A x B x D	2	3.063	<1	
A x C x D	2	68.375	2.62	
B x C x D	1	84.000	3.22	
A x B x C x D	2	21.750	<1	
error	120	26.078		

ERIC

Table 5

Mean Base-Rate and Mean Difference Score for Each Main Effect

Effect	Mean Base-Rate	Mean Difference Score
Grade Level		
First	20.69	-.44
Third	23.46	1.59
Fifth	27.73	1.59
Sex of _E_		
Male	23.79	.95
Female	24.12	.88
Bias Condition		
To Males	23.11	1.06
To Females	24.81	.76
Sex of _S_		
Male	24.01	.37
Female	23.90	1.46

Table 6

Mean Base-Rate Scores for the Bias Condition x Sex of Subject Interaction

Sex of Subject	Bias Condition Males	Bias Condition Females
Males	21.78	26.25
Females	24.44	23.36

ERIC

significant Bias Condition effect only for the boys ($F = 13.79$, $df = 1/120$, $p < .001$), but significant Sex of Subject effects for both Bias to Males ($F = 4.88$, $df = 1/120$, $p < .05$) and Bias to Females ($F = 5.76$, $df = 1/120$, $p < .05$).

Analysis of Difference Scores. The difference scores were subjected to a 3 (Grade Level) x 2 (Sex of Experimenter) x 2 (Bias Condition) x 2 (Sex of Subject) x 6 (Minutes) analysis of variance (see Table 7). The means for the main effects are presented in Table 5. The significant Grade Level effect reflected higher difference scores for the third- and fifth-graders than for the first-graders (see Table 5). Female Ss had higher difference scores than male Ss (see Table 5). These interactions are of limited interest however, in view of the significant Grade Level x Bias Condition x Sex of Subject interaction (see Table 8). Individual comparisons (Winer, 1962, p. 344) were conducted on the bias condition x sex of subject means separately for each grade level and tests of simple effects (Kirk, 1968, p. 289ff) were conducted on the grade level x sex of subject means for each bias condition. The individual comparisons revealed no significant Bias Condition or Sex of Subject effects at the first-grade level. At the third-grade level there was a significant Bias Condition effect ($F = 12.20$, $df = 1/120$, $p < .001$) for the males but not for the females. There were significant sex differences for both the Bias to Male ($F = 7.36$, $df = 1/120$, $p < .01$) and Bias to Female ($F = 4.22$, $df = 1/120$, $p < .05$) conditions. At the fifth-grade level the bias conditions were significantly different for both the male ($F = 7.72$, $df = 1/120$, $p < .01$) and female Ss ($F = 16.11$, $df = 1/120$, $p < .001$). The tests of simple effects revealed that for the Bias toward Males condition there was a significant age effect ($F = 10.43$, $df = 2/120$, $p < .001$) for the male Ss but not for the female Ss. In the Bias toward Females condition the age effect was

Table 7

Analysis of Variance of Difference Scores

Source	df	MS	F	P
Grade Level (A)	2	394.066	7.22	<.001
Sex of E (B)	1	1.260	<1	
Bias Condition (C)	1	19.260	<1	
Sex of S (D)	1	254.584	4.67	<.05
A x B	2	19.448	1	
A x C	2	81.816	1.50	
A x D	2	225.876	4.14	<.05
B x C	1	13.751	<1	
B x D	1	10.446	<1	
C x D	1	1641.760	30.10	<.001
A x B x C	2	6.689	<1	
A x B x D	2	78.300	1.43	
A x C x D	2	166.774	3.06	<.06
B x C x D	1	19.862	<1	
A x B x C x D	2	4.689	<1	
error	120	54.546	<1	
Minutes (E)	5	.874	<1	
A x E	10	12.205	2.48	<.01
B x E	5	8.841	1.80	
C x E	5	3.074	<1	
D x E	5	3.493	<1	
A x B x E	10	6.853	1.40	
A x C x E	10	6.297	1.28	
A x D x E	10	5.651	1.94	
B x C x E	5	4.660	<1	
B x D x E	5	2.659	<1	
C x D x E	5	11.008	2.24	<.06
A x B x C x E	10	3.480	<1	
A x B x D x E	10	4.914	1.00	
A x C x D x E	10	3.996	<1	
B x C x D x E	5	8.826	1.80	
A x B x C x D x E	10	1.891	<1	
error	600	4.911		

Table 8

Mean Difference Scores for the Grade Level x Bias Condition x Sex of Subject Interaction

Bias Condition	Sex of Subject	Grade Level First	Third	Fifth
Males	Male	-.03	3.94	1.78
	Female	-.56	.60	.64
Females	Male	-1.47	-.36	-1.64
	Female	.31	2.17	5.58

Table 9

Mean Difference Scores for the Grade Level x Minutes Interaction

Grade Level	Minutes 2	3	4	5	6	7
First	.12	-.21	-.75	-.46	-.19	-1.15
Third	1.02	1.79	1.33	1.83	1.88	1.67
Fifth	1.62	1.27	1.71	1.56	1.02	2.35

significant for both the males (F = 6.45, df = 2/120, p <.01) and the females (F = 18.86, df = 2/120, p <.001). In effect, both the individual comparisons and the tests of simple effects reveal grade level differences and sex differences in susceptibility to subtle cues emitted by the experimenters.

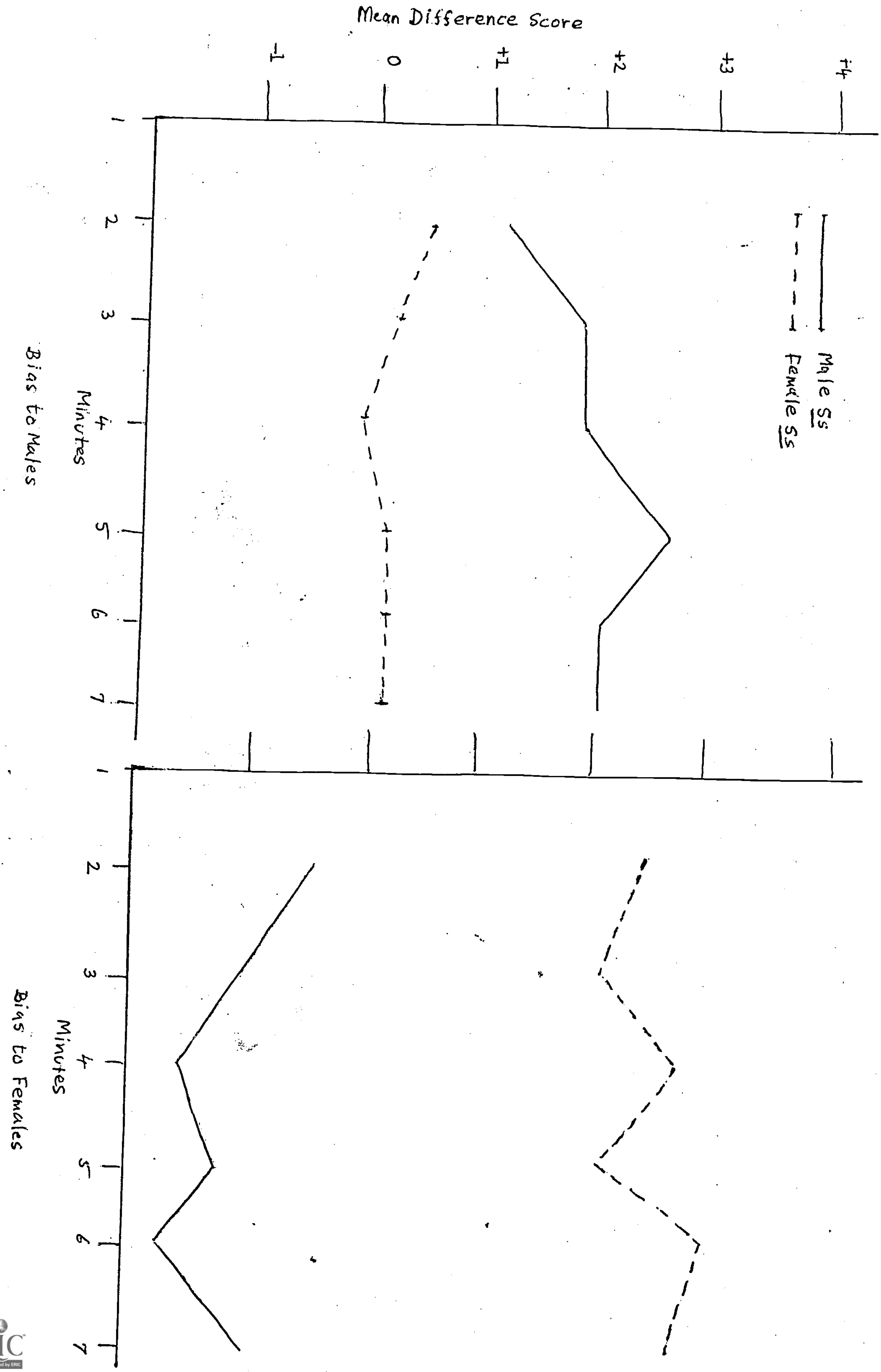
There were two within subjects comparisons which were significant. The means for the Grade Level x Minutes Interaction may be seen in Table 9. In general, the performance rates of the first-graders declined over time but the performance rates of the third- and fifth-graders increased over time and then remained relatively stable.

The Bias Condition x Sex of Subject x Minutes interaction approached the traditional p <.05 level of significance. The performance curves reflected by this effect are shown in Figure 1. Individual comparisons (Winer, 1962) of each pair of means for each minute revealed no significant differences between the sexes in the Bias to Males condition. Individual comparisons for the Bias to Females condition, however, revealed significant sex differences at each minute.

Discussion

Generally speaking, the experimenters did bias the performance of the children. The significant Bias Condition x Sex of Subject interaction in the analysis of the difference scores reveals differences in the predicted direction, i.e., boys performed at a higher rate than girls for experimenters biased toward males and girls performed at a higher rate than boys for experimenters biased toward girls.

The major finding with respect to the predictions was the Grade Level x Bias Condition x Sex of Subject interaction (Table 8) which revealed clear developmental trends in susceptibility to experimenter bias effects. At the first-grade level there were no significant Bias Condition effects for either the male or female Ss, although


Mean Difference Score
-1
0
+1
+2
+3
+4
Male Ss
Female Ss
Minutes
Bias to Males
Minutes
Bias to Females

the means were in the predicted directions. At the third-grade level the Bias Condition effect was significant for the males, the mean difference score was higher if the experimenter was biased toward males than females; for the females the Bias Condition effect was not significant although the means were in the predicted direction. At the fifth-grade level the Bias Condition effect was significant, for both the male and female subjects with the means in the predicted direction.

The above findings indicate clear developmental and sex of subject trends in susceptibility to experimenter-bias effects. Although the exact bases of these trends is difficult to elaborate at the present time it may be that as children become more developmentally mature they are better able to interpret the subtle cues emitted by the experimenter and tend to comply with the interpretation placed on the cues. The processes involved may be similar to those examined by Flavell (1968) in connection with children's role-taking and communication skills.

Experiment III

Adult Expectancy Effects: Self-generated versus Induced Expectancies

When evaluating adult expectancy effects there is but little evidence relating to the importance of the manner by which the adult acquires the expectancy for the to-be-produced outcome. Some of the available evidence (e.g., Bootzin, 1971) suggests that self-generated expectancies relate more to obtained bias than expectancies induced by the principal investigator. However, there is other evidence (e.g., Marcia, 1961; Marwit & Marcia, 1967) which does not support this position. Experiment III was aimed at assessing the importance of mode of development of adult expectancies vis-a-vis the effectiveness of adults to bias the simple motor performance of children.

Subjects

The subjects were 48 kindergarten children ($\overline{CA}$ = 5 yrs. 11 mo., SD = 8 mo.), half males and half females. The children attended a school serving primarily a lower-class neighborhood.

Experimenters

The experimenters were 12 male college students ($\overline{CA}$ = 19 yrs., 9 mo., SD = 13 mo.) enrolled in the introductory psychology course at Syracuse University. During the experiment each E tested two boys and two girls. Each E was randomly assigned to one bias condition and one induction condition.

Apparatus

The apparatus was identical to that used in Experiment II.

Procedure

Experimenter training. With the exception of the group training session the experimenter training was essentially identical to that of Experiment II. Each experimenter was trained individually. Experimenters assigned to the Induced Bias Condition were given the same statement as was given in Experiment II. Experimenters in the Self-generated Bias Condition were asked to predict whether boys or girls would drop the marbles faster. Each experimenter practiced the task administration procedures with a graduate student.

Experimental task. The procedures for the experimental task were identical to those employed in Experiment II.

Design

The design of Experiment III was a 2 (Induction Condition) x 2 (Bias Condition) x 2 (Sex of Subject) x 7 (Minutes) factorial design with six subjects in the smallest cell.

ERIC

Results

Dependent measures. As in Experiment II, the dependent measures were the base-rate of response and six difference scores, one for each minute of the experimental period of the task. The correlation between the base-rate and the average difference score was -.62 (N = 48, p <.01), indicating that approximately 38% of the variance in the difference scores is accounted for by the initial base-rate levels.

Analysis of base-rate scores. The base-rate scores were subjected to a 2 (Induction Condition) x 2 (Bias Condition) x 2 (Sex of Subject) analysis of variance (see Table 10). The mean base-rate scores for each main effect are presented in Table 11. As may be seen in Table 10 the only significant effect was the triple interaction involving Induction Condition, Bias Condition, and Sex of Subject. The means for this effect are presented in Table 12. Individual comparisons revealed the following: a) the only significant sex difference (F = 7.08, df = 1/40, p <.05) was in the Self-Generated Bias to Females Condition; b) the only significant Bias Condition difference was for the male Ss in the Self-Generated Induction Condition.

Analysis of Difference Scores. The difference scores were subjected to a 2 (Induction Condition) x 2 (Bias Condition) x 2 (Sex of Subject) x 6 (Minutes) analysis of variance with repeated measures on the last factor (see Table 13). The means for the between-subjects main effects are presented in Table 11. The only significant between-subjects effect was the Bias Condition x Sex of Subject interaction (see Table 14). Individual comparisons (Winer, 1962, p. 344) revealed significant Sex of Subject effects for Bias to Males (F = 6.48, df = 1/40, p <.05) and Bias to Females (F = 22.90, df = 1/40, p <.001) and significant Bias Condition effects for Male Ss (F = 10.36, df = 1/40, p <.01) and Female Ss (F = 16.92, df = 1/40, p <.001).

There were several significant within-subjects effects. The significant Minutes main effect (See Table 15) reflected a general decrease in rate of response during the experimental period of the task. The Induction Condition x Minutes interaction was significant (see Table 15) and reflected a general decrease in rate of response for Ss in the Induced condition and, generally, an increase and then decrease in rate of response for Ss in the Self-Generated condition. The Induction condition x Bias Condition x Minutes interaction was also significant (see Figure 2). Individual comparisons revealed that for the Induced Condition there were significant Bias Condition effects for minutes 4 ($F = 6.50$, $df = 1/200$, $p < .05$), 5 ($F = 6.03$, $df = 1/200$, $p < .05$), 6 ($F = 6.36$, $df = 1/200$, $p < .05$), and 7 ($F = 6.56$, $df = 1/200$, $p < .05$), but for the Self-Generated Condition the only significant Bias Condition effect was for minute 4 ($F = 4.27$, $df = 1/200$, $p < .05$).

Discussion

The major focus of Experiment III was to investigate the effects of mode of inducing expectations in the experimenter in the obtaining of experimenter-bias effects with children. Although the analysis of the difference scores revealed a significant Bias Condition x Sex of Subject interaction, indicating a significant experimenter-bias effect, this effect did not interact with Induction Condition. Induction Condition did interact with Minutes, and with Bias Condition and Minutes. However, these effects are not readily interpretable given current theorizing in the area. The data would appear to support the findings and theorizing of Marcia (1961) and Marwit and Marcia (1967), indicating no significant differences due to Induction Condition.

Table 10

Analysis of Variance of Base-Rate Scores

Source	df	MS	F	P
Induction Condition (A)	1	17.516	<1	
Bias Condition (B)	1	1.688	<1	
Sex of Subject (C)	1	17.520	<1	
A x B	1	46.023	2.44	
A x C	1	6.023	<1	
B x C	1	50.020	2.66	
A x B x C	1	88.023	4.67	<.05
error	40	18.838		

Table 11

Mean Base-Rate and Mean Difference Score for Each Main Effect

Effect	Mean Base-Rate	Mean Difference Score
Induction Condition		
Induced	17.92	-.75
Self-generated	16.71	.56
Bias Condition		
To Males	17.50	.43
To Females	17.12	.88
Sex of Subject		
Males	17.92	.10
Females	16.71	1.20

Table 12

Mean Base-Rates for the Induction Condition x Bias Condition x Sex of Subject Interaction

Induction Condition	Bias Condition	Sex of Subject Male	Female
Induced	To Males	19.67	18.50
	To Females	16.67	16.83
Self-Generated	To Males	14.50	17.33
	To Females	20.83	14.17

Table 13

Analysis of Variance of Difference Scores

Source	df	MS	F	P
Induction Condition (A)	1	2.722	<1	
Bias Condition (B)	1	14.222	<1	
Sex of Subject (C)	1	86.681	2.44	
A x B	1	102.722	2.90	
A x C	1	8.681	<1	
B x C	1	946.125	26.675	<.001
A x B x C	1	5.014	<1	
error	40	35.469		
Minutes (D)	5	28.431	6.60	<.001
A x D	5	10.631	2.47	<.05
B x D	5	.631	<1	
C x D	5	6.255	1.45	
A x B x D	5	9.797	2.28	<.06
A x C x D	5	.556	<1	
B x C x D	5	3.867	<1	
A x B x C x D	5	4.256	<1	
error	200	4.306		

ERIC

Table 14

Mean Difference Scores for Each Bias Condition x Sex of Subject Subgroup

Bias Condition	Sex of Subject Male	Female
To Males	1.69	-.83
To Females	-1.49	3.24

Table 15

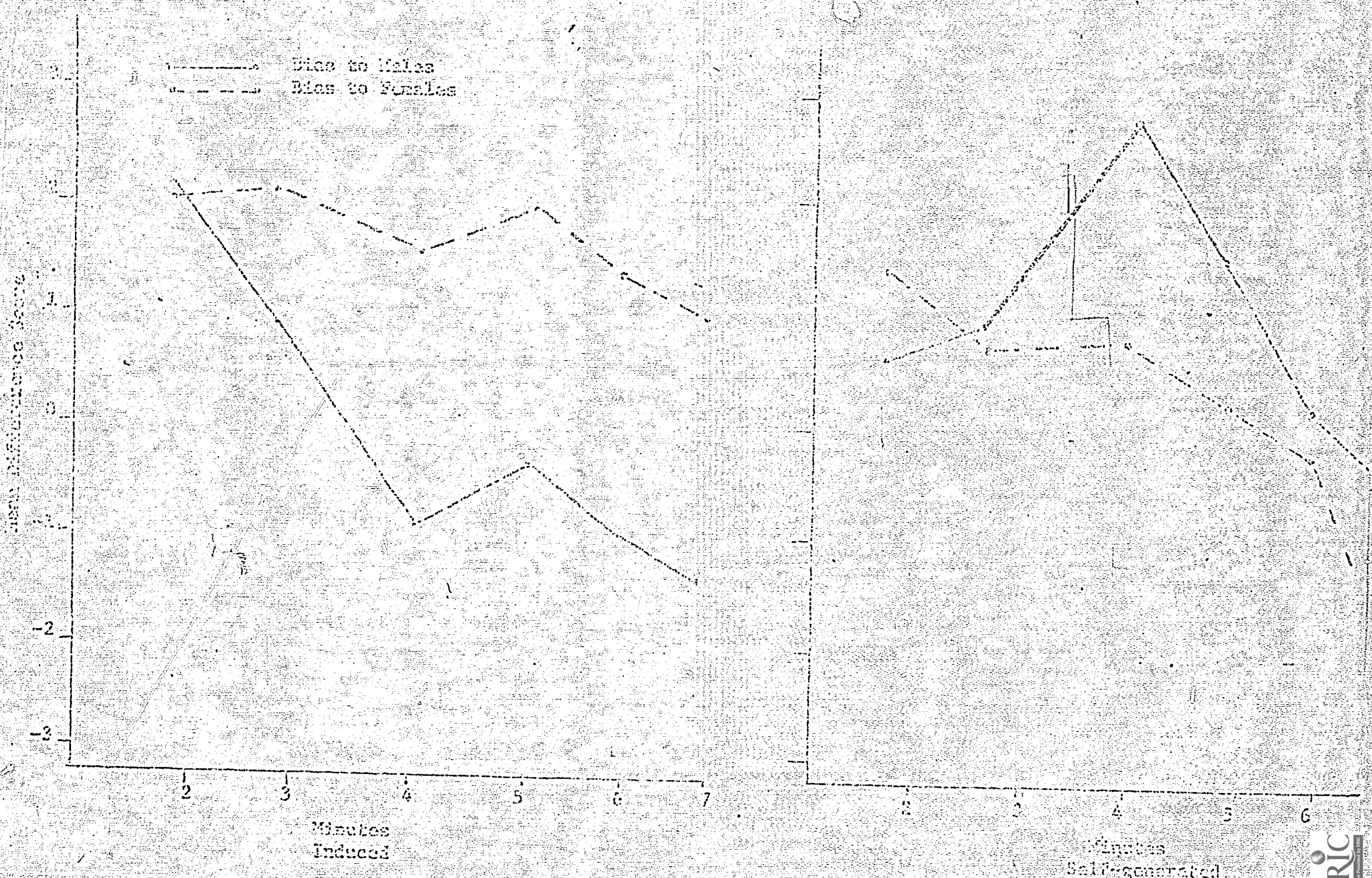
Mean Difference Scores for the Minutes x Induction Condition Interaction

Induction Condition	Minutes 2	3	4	5	6	7
Induced	2.04	1.42	.25	.79	.17	-.17
Self-generated	.88	.92	1.71	.79	.17	-1.12
Mean	1.46	1.17	.98	.79	.17	-.64

ERIC

Figure 7

Mean Difference Score for Each Induction Condition x Bias Condition Group Over Minutes



General conclusions from Experiments II & III:

Recall that in the study of teacher-bias and teacher-expectancy effects there were no Grade Level x Bias Condition interactions, indicating no developmental trends in teacher-bias or teacher-expectancy effects. The experimental study aimed at assessing developmental trends in experimenter-bias, however, did reveal a significant Grade Level x Bias Condition x Sex of Subject interaction. Clear developmental trends were discernible within this interaction. Older children evidenced a greater susceptibility to experimenter-bias effects than younger children. This divergence in findings is most likely due to the "group" nature of the teacher-bias experiment and the opportunity for single-subject interaction in the experimenter-bias study. That is, it may be more likely for developmental trends in susceptibility to adult influence to be evidenced in a one to one situation as opposed to a one to group situation.

Although the study of teacher-bias and teacher-expectancy effects revealed that teacher-expectancy but not teacher-bias effects related to SAT performance the findings of Experiment III were inconclusive with respect to differential effects of experimenter-bias and experimenter-expectancy in relation to simple motor performance. Again, it may be that bias and expectancy effects exert differential influences on performance depending upon the type of situational interaction, i.e., group or individual.

An alternative explanation for the divergence of findings must also be considered. It may be the case that, on the one hand, cognitive performance, such as is measured by achievement tests, is not susceptible to bias effects. On the other hand, motor performance may be influenced by such effects. Obviously, this is a question which must be answered by future research.

References

Anderson, D. F., & Rosenthal, R. Some effects of interpersonal expectancy and social interaction on institutionalized retarded children. Proceedings of the 76th Annual Convention, APA, 1968. Pp. 479-480.

Barber, X., & Silver, M. Fact, fiction and the experimenter bias effect. Psychological Bulletin Monograph Supplement, 1968, 70, 1-29. (a)

Barber, X., & Silver, M. Pitfalls in data analysis and interpretation: A reply to Rosenthal. Psychological Bulletin Monograph Supplement, 1968, 70, 48-62. (b)

Beez, W. V. Influence of biased psychological reports on teacher behavior and pupil performance. Proceedings of the 76th Annual Convention, APA, 1968. Pp. 605-606.

Bootzin, R. Induced and stated expectancy in experimenter bias. Proceedings of the 76th Annual Convention, APA, 1969, 4(1), 365-6.

Claiborn, W. L. Expectancy effects in the classroom: A failure to replicate. Journal of Educational Psychology, 1969, 60, 377-383.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70, 426-443.

Dusek, J. B. Experimenter-bias effects on the simple motor task performance of low- and high-test anxious boys and girls. Psychological Reports, 1972, 30, 107-114.

Dusek, J. B., & O'Connell, E. J. Teacher expectancy effects on the achievement test performance of elementary school children. Journal of Educational Psychology, 1973, in press.

Elashoff, J. D., & Snow, R. E. A case study in statistical inference: Reconsideration of the Rosenthal-Jacobson data on teacher expectancy. Technical report #15, Stanford Center for Research and Development in Teaching, School of Education, Stanford University, December, 1970.

Elashoff, J. D., & Snow, R. E. (Eds.) Pygmalion reconsidered. Worthington, Ohio: Charles A. Jones, 1971.

Evans, J., & Rosenthal, R. Interpersonal self-fulfilling prophecies: Further extrapolations from the laboratory to the classroom. Proceedings of the APA Convention, 1969, 371-372.

Flanagan, J. C. Test of general ability: technical report. Chicago: Science Research Associates, 1960.

Flavell, J. The development of role-taking and communication skills in children. New York: John Wiley & Sons, Inc., 1968.

Fleming, E. S., & Anttonen, R. G. Teacher expectancy as related to the academic and personal growth of primary-age children. Monographs of the Society for Research in Child Development, 1971, 36(50), Serial No. 145.

Friedman, N. The social nature of psychological research: The psychological experiment as a social interaction. New York: Basic Books, 1967.

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Jose, J., & Cody, J. Teacher-pupil interaction as it relates to attempted changes in teacher expectancy of academic ability and achievement. American Ed. Research Journal, 1971 8, 39-49.

Kirk, R. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Cole Publishing Co., 1968.

Marcia, J. E. The need for social approval, the condition of hypothesis-making, and their effects on unconscious experimenter bias. Unpublished master's thesis, Ohio State University, 1961.

Marwit, S. J., & Marcia, J. E. Tester bias and response to projective instruments. Journal of Consulting Psychology, 1967, 31, 253-258.

Meichanbaum, D., Bowers, K., & Ross, R. A behavioral analysis of teacher expectancy effect. Journal of Personality and Social Psychology, 1969, 13, 306-316.

Overall, J. E. Multiple covariance analysis by the general least squares regression method. Behavioral Science, 1972, 17, 313-320.

Overall, J. E., & Spiegel, D. K. Concerning least squares analysis of experimental data. Psychological Bulletin, 1969, 72, 311-322.

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

Rosenthal, R. Experimenter expectancy and the reassuring nature of the null hypothesis decision procedure. Psychological Bulletin Monograph Supplement, 1968, 70, 30-47.

Rosenthal, R. Empirical vs. decreed validation of clocks and tests. American Educational Research Journal, 1969, 6, 689-691. (a)

Rosenthal, R. Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal and R. Rosnow (Eds.), Artifact in behavioral research. New York: Academic Press, 1969. Pp. 182-277. (b)

Rosenthal, R. On the social psychology of the self-fulfilling prophecy: Further evidence for Pygmalion effects and their mediating mechanisms. Unpublished manuscript, 1973, Harvard University.

Rosenthal, R., & Jacobson, L. Pygmalion in the classroom. New York: Holt, 1968.

Rosenthal, R., & Rubin, D. B. Pygmalion reaffirmed. In J. D. Elashoff and R. E. Snow (Eds.), Pygmalion Reconsidered. Worthington, Ohio: Charles A. Jones, 1971. Pp. 139-155.

Rubovits, P., Maehr, M. Pygmalion analyzed: Toward an explanation of the Rosenthal-Jacobson findings. Journal of Personality and Social Psychology, 1971, 19, 197-203.

Snow, R. Unfinished Pygmalion. Contemporary Psychology, 1969, 14, 197-199.

Stevenson, H. W., & Fahel, L. The effect of social reinforcement on the performance of institutionalized and noninstitutionalized normal and feebleminded children. Journal of Personality, 1961, 29, 136-147.

Thorndike, R. L. Review of Pygmalion in the classroom. American Educational Research Journal, 1968, 5, 708-711.

Thorndike, R. L. You have to know how to tell time. American Educational Research Journal, 1969, 6, 692.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Footnotes

1. The project presented or reported herein was performed pursuant to a grant (No. OEG-2-71-0516) from the U.S. Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein, however, do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

2. The author is grateful to Mr. James McGee, Principal, and the teachers of the kindergarten through fifth-grades of Clinton Elementary School for their excellent cooperation and enthusiasm.

3. Analyses of SAT-1 scores revealed no differences between the children remaining available at the end of the first year or middle of the second year and those lost throughout the experiment.