DOCUMENT RESUME

ED 028 518 EA 002 116

By-Walberg, Herbert J.

A Model for Research on Instruction.

Pub Date Dec 68

Note-21p.; Preliminary draft (Dec. 1968) of paper presented at the Annual Meeting of the Amer. Educ. Res. Assn. (Los Angeles, Calif., Feb. 1969).

EDRS Price MF-$0.25 HC-$1.15

Descriptors-*Academic Aptitude, *Educational Environment, Educational Research, *Evaluation, Hypothesis Testing, *Instruction, *Learning Characteristics, Learning Theories, Models

Identifiers-Harvard Project Physics

To test the hypothesis that learning depends on the influence and interaction of three primary variables--instruction, learner aptitude, and learning environment--an explanatory model was tested during the 1967-68 school year in an experiment conducted with the cooperation of a simple random sample of 56 teachers drawn from the National Science Teachers Association of 17,000 secondary physics teachers. A number of pre- and posttests were administered to all 56 classes, with the experimental group of teachers using the Harvard Physics Project materials and the control group of teachers using their regular course materials. Findings of the study indicate the importance of recognizing not only instruction but also aptitude and learning environment as variables significantly affecting learning. (JK)

Read at American Educational Research Association Annual Meeting in Los Angeles February, 1969.

Preliminary Draft
December, 1968

A Model for Research on Instruction

Herbert J. Walberg

Harvard University

Evaluation of instruction rarely shows a clear superiority of one medium or method over another (Stephens, 1968). This paper is an attempt to show that much of the reliable variance in student performance is attributable to the aptitude of the learner and the environment of learning, leaving only a small part to be accounted for by instructional variables and perhaps by interactions between the three factors. Therefore, the design of research on instruction should include analyses of aptitude and environment and their interactions with instruction. To the extent that results of such designs are replicated and generalized, evaluation will contribute to basic educational research and to the formulation of verifiable principles and theories of instruction. Such a step is highly desirable since it is nearly impossible to evaluate all the many new media and methods of instruction.

Aptitude may be defined broadly as any characteristic of the individual or his experience that predicts any learning criterion. Measured intelligence, the most general predictor of school learning, typically accounts for 50 to 60 percent of the variance in school achievement. About 71 percent (the mean of estimates reviewed by Bloom, 1964) of the variance in intelligence itself can be attributed to heredity. An estimated 58 percent of the variance in intelligence can be predicted from measures of the home environment (Wolf, 1964). In contrast, parental socio-economic status, a much cruder and indirect measure of the home environment, typically accounts for about 16 percent of the variance in intelligence. These environmental sources of variance obviously overlap genetic sources of variance, e.g., bright parents have stimulating homes. The home environment also predicts about 64 percent of the variance on standardized achievement tests (Dave, 1964). From these and other data, Bloom (1964) has shown the importance of the child's early environment, especially during the first five years of life, in determining later achievement. Recent research has shown that measured environments of high school physics classes account for from 10 to 37 percent of the class variance in achievement, understanding, and interest criteria and from 9 to 16 percent after intelligence and other aptitudes are partialled out (Walberg, 1968). The learning environment may be defined as any stimulus, aside from the deliberate instruction in question, that predicts learning. The foregoing estimates of variance in learning accounted for by heredity and environment are considerably lower than could be accounted for if there were no error in measuring learning (see Bloom, 1964, for attenuation corrections and other considerations). In any case, there appears to be little variance in learning left to be accounted for by instructional differences after the effects of aptitude (a function of heredity and prior environment), the environment of learning during instruction, and measurement errors.

This conclusion is supported by Schutz's (1966) tabulation of the variance accounted for by experimental factors in research reported in the *Journal of Educational Psychology* and the *American Journal of Educational Research* in 1964. The median variance accounted for in studies of two group comparisons (t-tests) was 17 and the median for studies of three or more groups (F-tests) was 5. As Schutz points out, these figures have a strong upward bias since editors of these journals reject about 80 percent of submitted manuscripts and those rejected tend to be non-significant studies which account for little or no variance in learning criteria.

Thus, in formulating research on instruction, estimates of aptitudes and learning environments should be included. Such models would have several advantages: they would allow a determination of the relative effects of important factors; they would increase the precision and power of the analysis; and they would permit an investigation of the interactions of the factors. The remainder of this paper proposes such a model and reports some evaluation data which illustrate the potential value of the approach.

Equations 1 and 2 represent a model of the quasi-functional relationships between learning (L_h) and instruction (I_i), aptitude (A_j), the environment of

$$L_h = f(I,A,E) \qquad (1)$$

$$L_h = f_1(I_i)+f_2(A_j)+f_3(E_k)+f_4(I_iA_j)+f_5(I_iE_k)+f_6(A_jE_k)+f_7(I_iA_jE_k) \qquad (2)$$

learning (E_k), and (in Equation 2) their interactions. The functions of the first three terms, the main effects, are individually prominent in the research literature. They answer the following questions: Which instruction best promotes learning (f_1, sometimes called "summative evaluation")? Which students learn best (f_2, studies of prediction and selection)? Which environments best promote learning (f_3, stimulation and enrichment)? The first interaction, between instruction and aptitude (f_4), has been occasionally investigated in curriculum evaluations and is now being studied intensively by groups at Chicago and Stanford. To my knowledge, there are no studies of the second and last interactions (f_5, f_7), but Anderson is completing a dissertation at Harvard on the interaction of aptitude and environment (f_6). (Interactions are possible within each of the first three terms, of course, e.g., the interaction of media and teaching methods in the first term. The model may be expanded to include these more complex possibilities.) Many alternative models of instruction could be formulated, tested, and found useful. However, it has been shown that aptitude and environment are crucial; instruction must be included by definition. And variables from other models could be fitted into the terms of the present model, e.g., teacher characteristics and the time allowed for instruction in the first term, perseverance in the second term, and the ability to understand given instruction in the fourth term. The usefulness of any research model, of course, is in its comprehensiveness and capacities to be tested empirically and to generate research leading to new hypotheses and implications. Consider a test of the proposed model with some evaluation data from a national curriculum project.

Harvard Project Physics is a new high school course. Concerned about the steadily declining enrollments in physics (from 26 to 17 percent of high school students during the past 15 years), and the apparent failure of special groups to learn and appreciate physics, a group of educators and physicists attempted to develop a course that would not only appeal to potential scientists and technicians, but to those who might not be inclined to study physics, e.g., those with less ability and interest in science. As the course evolved, a number of explicit methods and media were produced to accomplish these ends. A text was written with a moderate density of physical concepts but with lower reading difficulty levels than other courses. Expanding on the idea of branching in programed instruction, the text included a number of cognitive modes of learning the same concept (or more advanced concepts), e.g., through philosophical explanations, mathematical formulations, historical narratives, and graphic representations. To arouse motivation in students uninterested in science and mathematics, tables, art works, and marginal notes were employed to illustrate concepts and their relations to the society and history of ideas. The text also allowed the student to pursue his own ideas and interests by referencing other instructional materials, a book of coordinated original readings, laboratory manuals and apparatus, programed instruction booklets, and film loops. So as not to overwhelm teachers or students with diversity but to allow as many choices as possible, teacher and student guides were produced which describe the organization of the materials and how best to use them for given purposes. It was also hoped that such materials and guides would compensate for poor facilities of some schools and the inadequacies of some physics teachers. Thus two main purposes of the Project are to develop methods and media suitable 1) for different students, especially scientifically disadvantaged and 2) for different conditions of instruction, especially those less than optimal.

In the proposed model for research on instruction, these purposes are represented by the instruction-aptitude and the instruction-environment interaction terms (f_4, f_5). It can be hypothesized that students with less aptitude and those in less desirable learning environments can still attain course objectives. To the extent this is confirmed, the course may be judged a success on these criteria. Examined below is an analysis on one criterion, gains in physics achievement, representing a small part of the total evaluation of the course and a preliminary test of the model.

Method

Sample, Testing Procedure, and Instruments

On the basis of anticipated costs, expected differences in courses, and standard deviations of variables in pilot studies, a sample size of 56 teachers was chosen. A simple random sample of teachers was drawn from the National Science Teachers Association of 17,000 physics teachers in the nation. All had agreed to teach Harvard Project Physics or their present course depending on our random assignment. Because special studies were planned of the Project teachers, about twice as many were assigned to teach the new course. The Project teachers were brought to Harvard for training during the summer. To minimize the so-called "Hawthorne effect", the control teachers were also brought in for a few days, given a cordial explanation of their role, but received no special instruction.

The teachers ~~were~~ administered a series of pretests, midtests, and posttests during the first two weeks in the fall term, in December, and the last two weeks of the spring term. Using random data collection (Walberg and Welch, 1967) random halves of each group took different tests simultaneously, a procedure which increases the number of instruments administered and decreases testing time for each student. However, biographical data were collected on all students in March.

The criterion instrument, the Physics Achievement Test, a 36-item, locally-constructed test, was given as a pretest and posttest. Items were written and chosen to represent general physics concepts. The Kuder-Richardson Formula 20 reliability of the Physics Achievement Test is .77 and the regression-adjusted gain score reliability using Traub's (1967) formula is .42. Plots of the gain scores and tests for skewness and kurtosis revealed no departure from the normal distribution. Since the test had been given to random halves of classes on the two occasions, a quarter of the sample of students (N=499) had taken it twice, and were therefore included in the sub-sample. Some workers hypothesize special sensitization to treatments may occur in experimental groups that are given the same pretest and posttest. However, the few empirical studies of the alleged effect contradict this hypothesis. Lana (1960) found that pretesting enhanced learning in experimental and control groups to a minor extent in a 12-day learning experiment, but the enhancement did not favor one group over the other. The sensitization effect is of little concern here since there was a long time period, 8 months, between testings.

A total of 31 aptitude and 17 environment measures thought to influence learning were selected for analysis. Among the pretests were measures of science understanding and interest in physical science. The Test on Understanding Science (Cooley and Klopfer, 1961) is a 60-item multiple-choice test on the nature of science, science as an institution, and scientists as people (see Table 3 for reliabilities of all measures). The sub-scale of Academic Interest Measure used (Halpern, 1965) consists of a number of items describing physical science activities, and the respondent is asked to indicate which he would like to do. His affirmative answers are summed for the sub-score.

In December the following instruments were administered: the Henmon-Nelson Intelligence Test (Lamke, Nelson, and Kelso, 1960), a selection of personality scales from those of Rokeach (1960) and Edwards (1959), and the Learning Environment Inventory (Walberg and Anderson, 1967, unpublished). The environment items describe characteristics of high school classes. The student expresses his agreement or disagreement on a 4-point scale. The mean of the items for each 14 sub-scores is calculated for student, and the mean of the sub-scores of the class is calculated for a group estimate of the learning environment. Prior work had shown that two weighted combinations of the scores, termed "Cognitive Press" and "Non-Cognitive Press," predict these two kinds of learning. They were also included in the analysis as was the class size of each student.

Nineteen multiple-choice items from the Biographical Inventory (Taylor and Ellison, 1967) were selected because they tap characteristics and experiences that may be related to learning (see Table 3). Lastly, the year in school was obtained from student questionnaire data and included in the analysis.

Analysis

Frequency distributions of the independent variables (except the biographical items) were used to choose cutting points at approximately the 20th and 80th percentiles of the total sample. The biographical items were roughly dichotomized at the 50th percentile. About 85 percent of the sample were seniors, so grade in school was split between seniors and others.

A preliminary least-squares analysis of variance of the course effects on boys and girls (see Tables 1 and 2) revealed that Project students gained significantly more (p less than .01) than students in other courses, that girls

(Tables 1 and 2 about here)

gained more than boys, and that there was no interaction between sex and course. More crucial for the planned analysis is that there were about twice as many girls as boys in the sample (reflecting national trends). Because of the grossly unequal cell sizes and because prior work (Walberg, 1969) had shown sharp differences in boys' and girls' reactions to physics, it was decided to analyze the groups separately.

Since the instruction-aptitude and instruction-environment interactions were of primary interest, 47 analyses of variance on boys and girls separately were computed, a large number but limited in terms of the model and the data available. These analyses were for the two course effects by three levels of environment and aptitude, except for the biographical items which had two levels. Each of the 31 aptitude and 16 environment measures were tested for significance by themselves and in their interaction with course. There were minor departures from proportional cell sizes, and a least-squares analysis for non-orthogonal designs, following Bock's (1963) algorythms, was employed. The effects are confounded in this kind of data, the estimates of effects and their significance levels are conservative for terms entering the model after other correlated effects. Hence, the course effect was entered first since the preliminary analysis showed it to be significant; the aptitude and environment terms were entered second in each analysis because they were of secondary interest; and the interaction terms were entered third as is recommended (Bock, 1963), and also because they were hypothesized.

Results and Discussion

Before turning to the hypothesized interactions, consider the effects of aptitudes and environments by themselves (i.e., without classifying students by course). The mean achievement gains for the significant trichotomized independent variables showed no curvilinear trends, i.e., for middle aptitude or environment groups to be higher or lower than both high and low groups. Hence a summary of the linear trends is displayed in the tabulation of correlations of the aptitude and environment score levels with achievement gains (Table 3). IQ is the only

(Table 3 about here)

aptitude that significantly predicts achievement gains for boys and girls. Many more of the other variables predicted for boys than for girls. The pattern for boys suggests an intellectual achievement propensity: scoring high on the science understanding test, getting superior marks, questioning teachers, entering science contests, following projects through, and planning college. Also, boys reporting high creative imagination and scoring low on Authoritarianism tended to gain more. The pattern is not as clear for girls who gained more: several bookcases in the home, entering music contests, and a low need for order.

The pattern of environment variables that predict gains for both boys and girls suggest intellectual challenge and group cohesiveness: high scores on Difficulty and Cognitive Press, low scores on Friction. The boys gained less in environments perceived as high on Favoritism and Disorganization.

The smaller sample size for girls may account for the fewer aptitude and environment measures that significantly predict their learning. However, recent research on high school physics students suggests that girls' docility and conforming behavior in school enables them to succeed more in science classes (though not in science careers) almost irrespective of their abilities, interests, and classroom conditions (Walberg, 1969).

Figures 1-3 show the statistically significant interactions (p less than .05) of central concern for evaluation of Project Physics, its capacity to promote

(Figures 1-3 about here)

learning in different students and learning environments. Figure 1 shows that "promising" boys in Project Physics and other courses all gain in physics achievement as much as would be predicted from their initial knowledge of physics. However, less promising boys in the Project gain as much as or better than would be expected, whereas the gains for less promising students in other courses are considerably lower--for those with low scores on the physics achievement and science understanding pretests, sophomores and juniors, and those who reported having less creative imaginations and fewer books in their homes. Project Physics also moderates the relationship between the most potent aptitude, IQ, and gains: the achievement of boys with less of this aptitude falls off more sharply in other courses than Project boys.

Similar results may be found in Figure 2 for boys in different learning environments. Project boys tend to gain more than predicted under most classroom environments whereas boys in other courses gain less than would be expected under several conditions: when the class is perceived as intimate, diverse, formal, moderately satisfying, and paced slowly.

There were only two significant course interactions for girls (see Figure 3). The first shows that girls in Project Physics gain more than predicted irrespective of year in school. Girls in other courses gain less if they are seniors and much better than expected if they are juniors or sophomores. Actually the latter group is very small (4 of the 115 subjects entering this particular analysis) since about 85 percent of the total sample were seniors, and the control group is about half the size of the Project group.

The other significant interaction for girls is complex and concerned Favoritism, the tendency for one student to be above another in the classroom hierarchy. Project girls gain more when the class is perceived as having low or moderate levels of favoritism; they gain less when the class is seen as having high levels of favoritism. Girls in other courses gain much less when the class is seen as having moderate levels of favoritism.

Thus it appears from the foregoing analysis of one criterion, Project Physics seems to be generally attaining two of its important objectives: enabling even students of relatively low aptitude to learn physics and enabling students in general to learn in class environments that hamper students in other courses. It must be admitted, however, that the present sample is a fairly bright group; the mean IQ is 116, reflecting the national trend for college-bound seniors to take physics. A national random sample was necessary to make inferences to the national population, but the Project staff are now making clinical studies of classes in urban slums to extend the evaluation to groups of lower aptitude than could be examined in the main study. It must also be admitted that the present evaluation is "molar" rather than "molecular;" it can only be concluded that providing many modes and methods of learning physics and guides for choosing among the options seems to promote learning for a wider variety of students and in different environments. It has not been possible to examine the "fine structure" of instruction and investigate the individual options and their interactions with aptitudes and environments. Comprehensive studies of this order of complexity are obviously difficult and time-consuming, certainly beyond present resources if not ambitions.

Summary and Conclusion

Aside from instruction, learning may be said to depend on two major classes of variables: aptitudes and learning environments. An aptitude may be defined broadly and operationally as any characteristic of the individual or his experience prior to instruction that predicts learning. The learning environment is any stimulus present in the environment, aside from instruction, that predicts learning. A model of research on instruction is proposed that includes instruction by definition, aptitudes, the learning environments, and their interactions. An example of the potential usefulness of the model is taken from the evaluation of Harvard Project Physics. Students in the course appear to succeed in gaining physics knowledge to some extent irrespective of a number of aptitudes and environments whereas students in other courses do not succeed when either their aptitudes or learning environments are less than optimal. The model points to a number of other relationships worth investigating.*

*The ideas in this paper have benefited from communications with a number of correspondents, colleagues, and students. I thank them for their kindness and absolve them of any responsibility for errors in the manuscript.

Table 1

Analysis of Variance for Course Effects on Physics Achievement Gains for Boys and Girls*

Source of Variance	df	Mean Square	F-Ratio
Course	1	341.56	13.72
Sex	1	111.37	4.47
Interaction	1	10.83	.44
Error	496	24.90	
Total	499		

Table 2

Gains in Physics Achievement by Course for Boys and Girls**

	Boys			Girls		
	M	SD	N	M	SD	N
Project Physics	.07	.70	250	.19	.64	80
Other	-.19	.72	127	.02	.74	42

*Note -- F-tests for Course and Sex are significant at the .01 level.

**Note -- The scores are expressed as the residual from prediction.

Table 3

Correlations of Aptitude and Environment Measures with Gains in Physics Achievement

	Gains	
	Boys	Girls
Pretest and Intelligence		
Physics Achievement (77)	2X	-2
Science Understanding(76)	22X**	16
Physical Science Interest(91)	3	6
IQ(91)	28X**	29**
Biographical Items		
Several bookcases in home	5X	15*
Much non-school reading	7	3
Much non-school study	4	-5
Much opportunity for creativity	5	-6
Often discuss occupational choice	-3	5
Like school	7	-9
Often question teachers	12*	0
Superior marks	18**	11
High persistence	0	-12
Always follow through projects	10*	4
Importance of intelligence	8	15
High creative imagination	12X**	-6
Often lead groups	-7	-8
Entered visual art contests	0	-2
Entered science contests	15**	7
Entered entertainment contests	-1	0
Entered music contests	7	20*
Entered writing contests	5	11
Senior year in high school	-6X	1X
Planning college	13**	1
Personality		
Dogmatism(65)	-1	-6
Authoritarianism(69)	-15*	-15
Rigidity(68)	1	-16
Achievement(78)	10	1
Order(82)	6	-27*
Affiliation(84)	6	1
Change(75)	12	-10

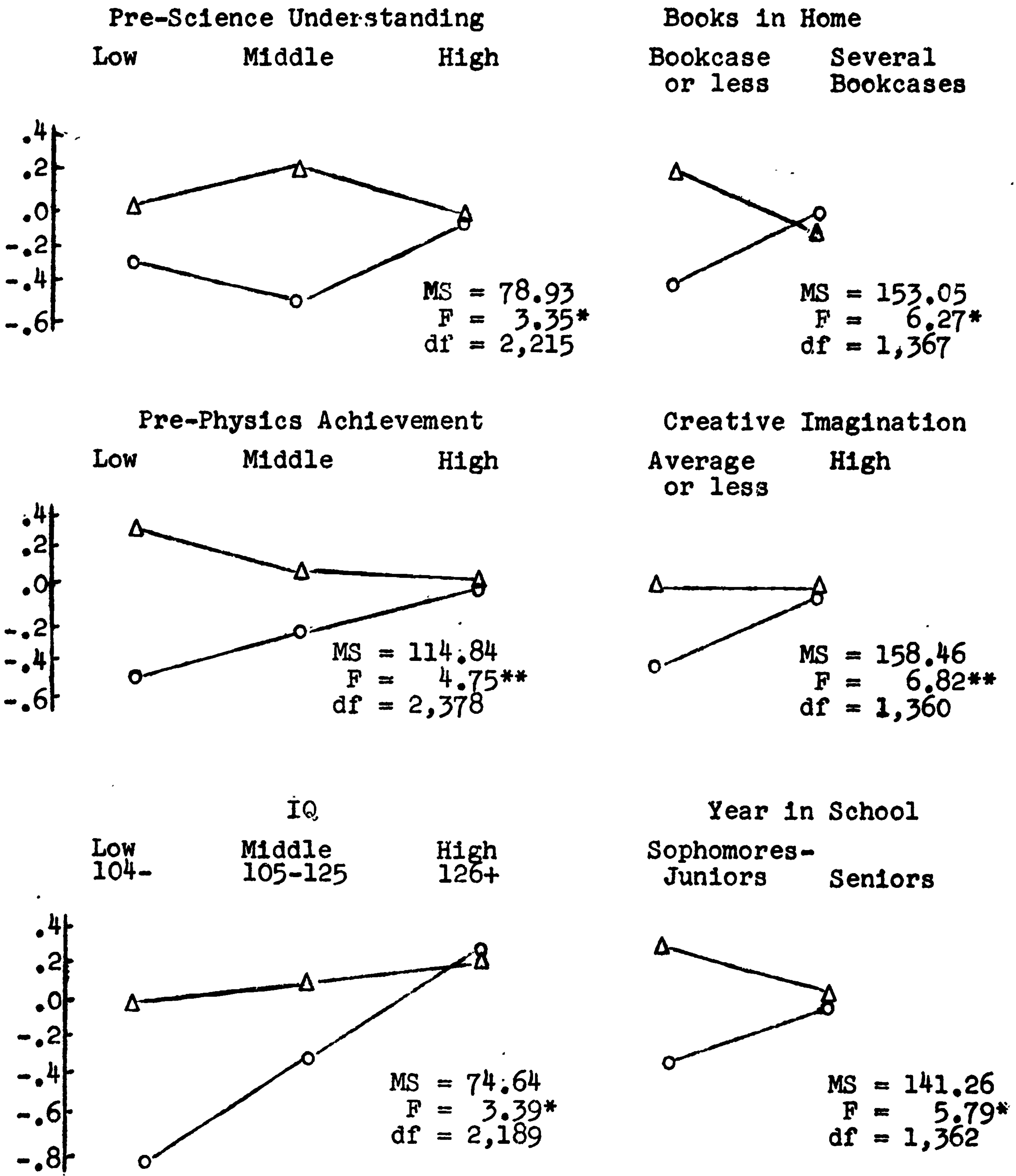
Table 3
(Continued)

	Gains	
	Boys	Girls
Learning Environment		
Intimacy(78)	-2X	-1
Friction(78)	-13**	-16*
Cliqueness(74)	-13**	-13
Satisfaction(80)	9X	7
Pace(77)	-1X	10
Difficulty(66)	13**	16*
Apathy(83)	3	4
Favoritism(77)	-9*	-3X
Formality(64)	4X	7
Direction(86)	0	-1
Democratic(67)	4	3
Disorganization(81)	-9*	-4
Diversity(58)	0X	-3
Cognitive Press	19**	18*
Non-Cognitive Press	-8	5
Class size	1	-10

Note -- Decimals omitted; read correlations in hundredths. Those significant at the .05 and .01 levels as main effects in the analyses of variance are indicated with one and two asterisks, respectively. Variables which interacted significantly with instruction for a given sex group are indicated with an "X". Reliabilities (either Kuder-Richardson Formula 20 or Cronbach alpha) are given in parentheses.

Figure 1

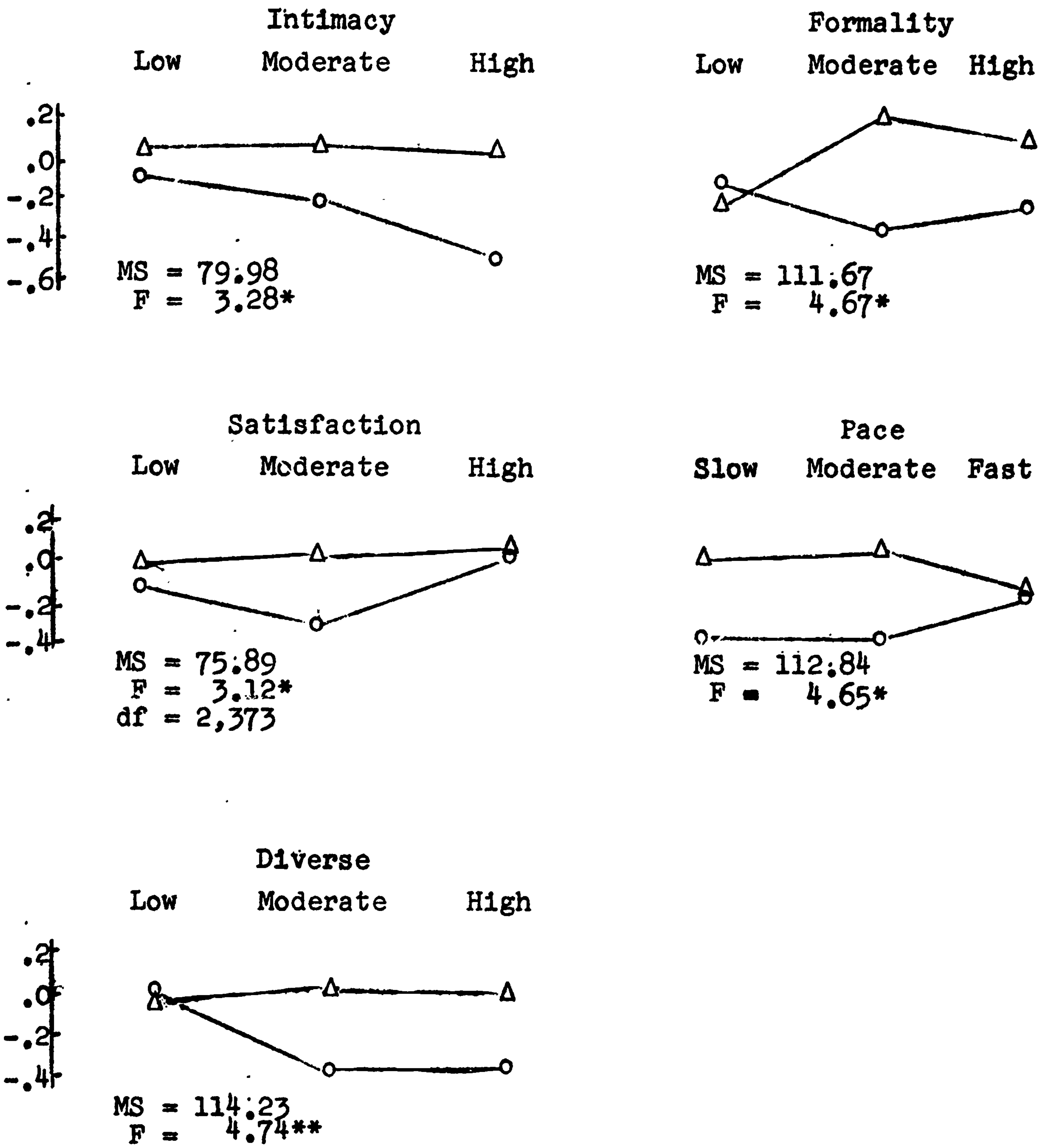
Mean Residual Gains for Boys in Project Physics (Δ) and Other Courses(o)



Note -- The mean squares, F-ratios, and degrees of freedom are for the interaction of course effects with each aptitude; F-ratios significant at the .05 and .01 levels are indicated with one and two asterisks, respectively.

Figure 2

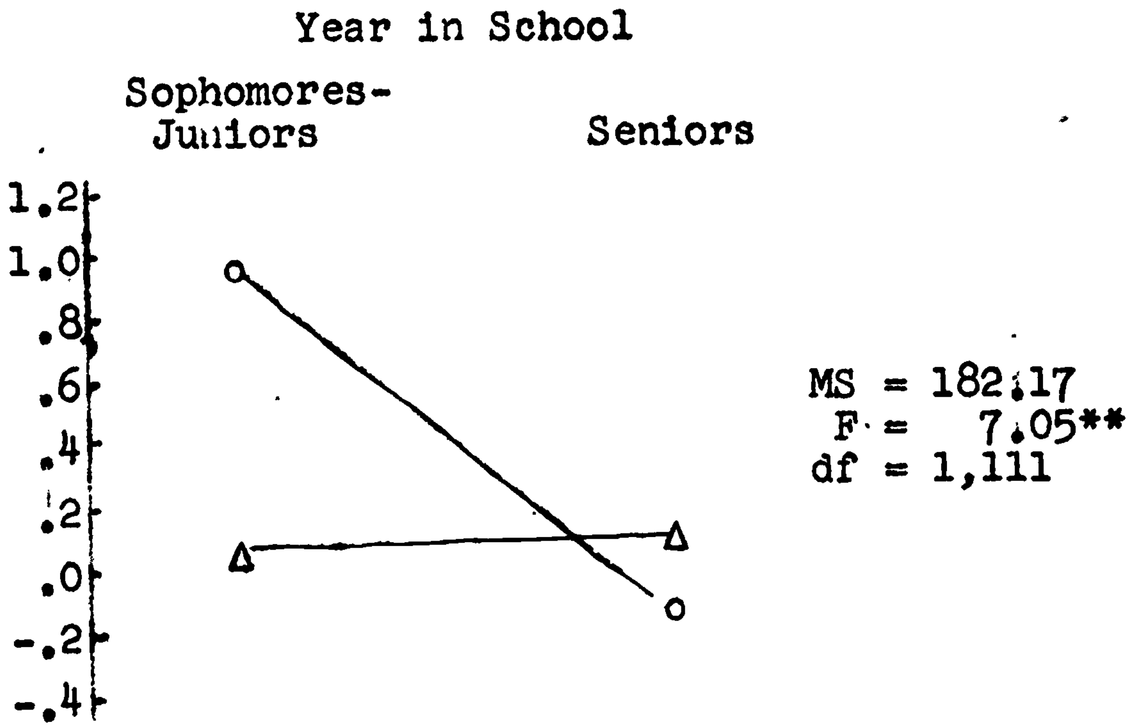
Mean Residual Gains for Boys in Project Physics(Δ) and Other Courses(o)

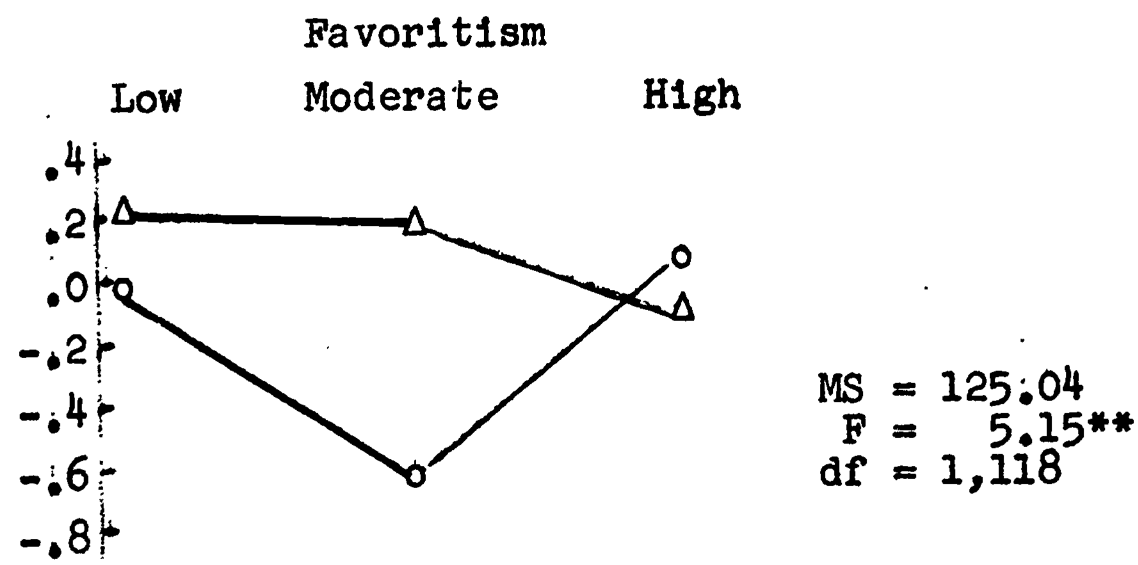


Note -- The mean squares and F-ratios are for the interaction of course effects with each environment measure. F-ratios significant at the .05 and .01 levels are indicated with 1 and 2 asterisks, respectively.

Figure 3

Mean Residual Gains for Girls in Project Physics(Δ) and Other Courses(o)





Note -- The mean squares, F-ratios, and degrees of freedom are for the interaction of course effects with each factor; both F-ratios are significant and indicated with 2 asterisks.

References

Bloom, Benjamin S., Stability and Change in Human Characteristics. New York: John Wiley & Sons, Inc., 1964.

Bock, R. Darrell, "Programming Univariate and Multivariate Analysis of Variance," Technometrics, 5:95-117, February, 1963.

Cooley, William W. and Klopfer, Leo, Manual for the Test on Understanding Science. Princeton, N.J.: Educational Testing Service, 1961, 12 pp.

Dave, R. H., "The Identification and Measurement of Environmental Process Variables that are Related to Achievement." Unpublished Ph.D. dissertation, University of Chicago, 1963.

Edwards, Allen L., Edwards Personal Preference Schedule. New York: Psychological Association, 1959.

Halpern, Gerald, Scale Properties of the Interest Index. Princeton, N.J.: Educational Testing Service, 1965, 23 pp.

Lamke, Tom A., Nelson, M. J. and Kelso, Paul C., Henmon-Nelson Tests of Mental Ability; Revised Edition. New York: Houghton Mifflin Co., 1960.

Lana, Robert E. and King, David J., "Learning Factors as Determiners of Pretest Sensitization," Journal of Applied Psychology, 44, 1960, 189-191.

Rokeach, Milton, The Open and Closed Mind. New York: Basic Books, 1960, 447 pp.

Schutz, Richard E., "The Control of 'Error' in Educational Experimentation." School Review, 150-158, Summer, 1966.

Taylor, Calvin W. and Ellison, Robert L., "Biographical Predictors of Scientific Performances," Science, 155:1075-1080, March, 1967.

Walberg, Herbert J. and Welch, Wayne W., "A New Dimension of Randomization in Experimental Curriculum Evaluation." School Review, 75:187-196, Summer, 1967.

Walberg, Herbert J., "The Social Environment as a Mediator of Classroom Learning." Submitted to the Journal of Educational Psychology, 1968.

Walberg, Herbert J., "Physics, Femininity, and Creativity." Developmental Psychology, 1969, in press.

Wolf, Richard M., "The Identification and Measurement of Environmental Process Variables Related to Intelligence." Unpublished Ph.D. dissertation, University of Chicago, 1964.

Stephens, J. M., The Process of Schooling: A Psychological Examination. New York: Holt, Rinehart and Winston, 1967.