DOCUMENT RESUME

ED 065 605 TM 001 875

AUTHOR Block, James H.
TITLE Student Evaluation: Toward the Setting of Mastery Performance Standards.
PUB DATE 4 Apr 72
NOTE 28p.; Paper presented at the annual meeting of the AERA (Chicago, Ill., April 4, 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Academic Achievement; *Cognitive Processes; Data Collection; Evaluation Techniques; *Feasibility Studies; Learning Processes; Models; Objectives; *Performance Criteria; Standards; *Student Evaluation; Task Performance; Test Results

ABSTRACT

When the task of evaluating student learning is carefully considered, two major problems emerge. One is the gathering of the most appropriate and precise evidence possible about the learning. The other is the setting of performance standards against which this evidence may be weighed and the adequacy of each student's learning judged. This paper has focused on the problem of setting performance standards for use in strategies for mastery learning. The paper began with the argument that a key variable in the design of these strategies are the mastery performance standards which students are helped to attain throughout their instruction. It was pointed out that presently there are no procedures for setting such standards. Next, an attempt was made to formulate one such procedure. The approach developed utilizes students' future learning, i.e., their scores on a set of desired, end-of-instruction learning outcomes, as a criterion for determining the mastery performance level which students must attain at any stage in their instruction. Finally, the paper reported an experiment designed to explore the feasibility of the approach proposed. The experiment was designed to test the assumption that the performance standard which a student attains over each segment of his instruction has important implications for his realization of the desired, end-of-instruction learning outcomes. In general, the experiment's results confirmed the assumption tested. (Author/CK)

4.1

STUDENT EVALUATION: TOWARD THE SETTING OF MASTERY PERFORMANCE STANDARDS

James H. Block

University of California, Santa Barbara

AERA

Chicago, Illinois

April 4, 1972

Although the problem of setting performance standards is old, perhaps never has it assumed greater importance than in the design of mastery learning strategies. These strategies are designed on the assumption that attainment of particular standards throughout the instructional process will help a maximum number of students reach desired end-of-instruction learning outcomes. Without procedures for selecting standards whose maintenance produces the desired outcomes, therefore, these strategies cannot be consistently well-designed.

Presently there are no such procedures. In part this is due to the type of standard which must be set. While there are well-developed procedures for setting relative (e.g., Angoff, 1971) or absolute (e.g., Nedelsky, 1954) standards for use in interpreting scores on norm-referenced tests, mastery performance standards must be absolute standards for use in interpreting scores on criterion-referenced tests (Block, 1971; Bormuth, 1971).[1]

In larger part, though, the lack of sound procedures for setting mastery performance standards is due to how these standards must be set. As Bormuth (1971) has argued the setting of mastery performance standards requires rational techniques which are capable of yielding standards whose superiority

[1] For purposes of this paper, a norm-referenced test may be defined as an instrument designed to indicate how well the student has learned a given segment of instruction relative to his peers while a criterion-referenced instrument may be defined as one designed to indicate what the student has or has not learned from the segment.

as indices of the adequacy of a student's learning can be defended both logically and empirically.

But present techniques for setting mastery standards are essentially irrational and yield indefensible standards. For example, perfect performance, i.e., a perfect test score over a set of instructional objectives, is often set as a mastery standard. Yet data from both laboratory learning research (Bormuth, 1971) and studies which employ perfect performance as their mastery standard (e.g., Sherman, 1967) suggest that perfect performance is unrealistic to expect and prohibitively expensive to attain. Or to take another example, mastery standards are also often set in terms of achievement of some fixed proportion, say 85 percent, of a set of objectives. But why this proportion is a more meaningful index of mastery than, say, achievement of 80 or 90 percent of the objectives is rarely explained or defended.

If the current trend toward increased use of mastery learning strategies continues, then the need for some better approaches to the problem of setting mastery performance standards is obvious. This paper attempts to begin to fulfill this need by (a) formulating a general approach for the setting of defensible mastery standards and (b) testing the approach's feasibility.

An Approach for Setting Mastery Performance Standards

According to Cahen (1970) one way to assess the learning outcomes of any instructional segment is to examine how well that segment has prepared the student for future learning. This idea has major implications for the setting of mastery performance standards for two reasons. First, in mastery strategies the major outcome over any instructional segment is the performance level to which each student learns. And second, these strategies are explicitly designed such that the level to which students are helped to learn should maximize each student's likelihood of attaining a set of desired, end-of-instruction learning outcomes.

In a mastery learning context, therefore, Cahen's notion can be translated as follows: one way to assess mastery over any segment of instruction is to examine how well the attainment of various performance levels prepares students for attaining the desired learning outcomes. That level which best prepares students vis-à-vis these outcomes can then be selected as one's mastery standard. For example, one could set a mastery performance standard for a two-unit instructional sequence where achievement and interest were the desired learning outcomes by determining that performance level whose attainment best maximized these outcomes.

Three major steps must be taken to implement this approach. First, the learning outcomes to be maximized by the instruction's completion must be characterized by a set of defensible learning criteria. This entails the creation of some practical (Schwab, 1968) methodology for making justifiable value-judgments. The method must be practical because we have no comprehensive theory of the outcomes of instruction (Bormuth, 1971; Gagné, 1970) and hence no theory which might guide one in selecting among existent criteria or generating new criteria for representing the learning. The method must make value-judgments, i.e., the selection of some subset of criteria from the range of possibilities available according to some priorities, because the choice of criteria represents essentially a value or values judgment (Scriven, 1967; Messick, 1970).

Next, some model for interrelating, weighting and combining scores over the various criteria must be developed (Bormuth, 1971). This model has two functions. First, it attempts to capture the wholeness and the complexity of the learning to be maximized. Second, it provides a decision function for selecting that standard which best maximizes the learning in cases where the attainment of different performance levels maximizes scores on different criteria. Suppose, for example, that the learning to be maximized is represented by two criteria, achievement and interest, and that the attainment of one performance level maximizes future achievement while the attainment of another maximizes future interest. Depending upon one's model, the standard would be set either closer to the first performance level or closer to the second. If achievement plays a far larger role than interest in the model, then the standard would be set closer to the first level. But if interest plays a far larger role than achievment, then the converse would be true.

Finally, having defined the learning to be maximized by a set of defensible criteria and a model which incorporates these criteria, maximal learning must be clearly defined. Here a statistical technique must be selected for estimating future learning (i.e., estimating scores from the model) as a function of the performance level to which the unit or units over which the standard will be set are learned. Least-squares (e.g., regression), Bayesian or other estimation procedures might be used. That level which yields the greatest estimated future learning can then readily selected as one's mastery standard.

Advantages of the Proposed Approach

While this approach to the setting of mastery standards is neither as simple or as expedient as its predecessors, it does have some powerful advantages. First, it introduces a heretofore missing element of objectivity into the process of setting mastery standards. From the choice of learning criteria to the selection of some statistical estimation technique, one is forced to be explicit about the decision processes by which he arrives at his standards. Thereby he opens his standard setting process to scrutiny and challenge by other individuals and enables these individuals, if they so choose, to independently verify his standard through replication.

Second, the procedure yields standards which have clear meaning for student learning. It enables one to set a standard whose attainment should lead to greater future learning than would have the attainment of any other standard.

And third, it allows one to optimally design his mastery learning strategy. As Glaser and Nitko (1971) have pointed out, "Instruction proceeds as a function of the relationship among measures of student performance, available instructional alternatives and learning criteria that are chosen to be optimized." The approach developed here establishes a clear relationship among these variables. It forces one to select particular learning criteria to be optimized and to set standards whose attainment will clearly optimize this learning. One can then select from the available instructional alternatives the particular design which ensures that the standard selected will be maintained and hence that the desired end-of-instruction learning outcomes will be reached.

A Feasibility Study

The mastery standard setting approach outlined above rests upon many assumptions. But perhaps the most basic of these assumptions is that the performance standard which a student is helped to attain over each segment of his instruction has important implications for his realization of the desired end-of-instruction learning outcomes. This segment of the paper reports the results of a study designed to test this assumption and, hence, the approach's feasibility.

Method

Subjects: Ninety-one eighth graders from a lower-middle class suburb near Portland, Oregon formed the sample.

Learning Materials: Matrix algebra was selected as the subject matter for this study because it best fit the following requirements. First, it is taught sequentially, that is, each segment of instruction builds upon the prior segment. If the performance level which a student is helped to attain over each segment of his instruction does influence his future learning, then it seemed reasonable that these effects would be clearest in sequentially taught subjects. Second, the algebra was sufficiently relevant to the student's prior learning so that it would not be perceived as being so difficult that only a few could learn it or so easy that it would be sheer busywork. Finally, the algebra was sufficiently esoteric to ensure no spill over of any negative experimental effects into the students other school work.

Three programmed units in elementary matrix algebra were developed from a textbook developed by Bhushan (1968): Unit I - The definition and some properties of matrices; Unit II - Special types of matrices and the rules of matrix equality; and Unit III - the process and rules of matrix addition and the process of matrix subtraction. Each unit was constructed so that most students would learn only about 50 percent of the material from the text alone.

Learning Criteria: Common school learning criteria were used to represent the future learning to be maximized, i.e., the goals of the instruction. The first criterion was Achievement. This is the criterion most often used by schools to measure a student's learning. It indicates his acquistion of the intellectual skills (content and mental processes) taught and also serves much like aptitude and intelligence measures as an index of general learning capacity.

Achievement, however, may be thought of a indexing only the level to which a student has learned. Bloom (1968), drawing on the work of Carroll (1963), has suggested that the rate at which a pupil learns to a given level or the level to which he learns in a given amount of time are inter-changeable learning criteria. Hence, although some have challenged the utility of rate measures (e.g., see Cronbach and Snow, 1969), the Time Needed to Learn was chosen as a second criterion.

A third criterion was Transfer where transfer was defined as the application (Bloom, 1956) of cognitive skills achieved under one set of conditions to the solution of related new problems. This criterion was selected since neither a high level of achievement or a quick learning rate guaranteed that the student would be able to apply the skills acquired by one point in time at a future point. Much of school learning, though, is cumulative precisely in the sense that what is learned at one point must be applied at some later point to facilitate new learning (Gagne, 1965) or to solve new problems (Brownell, 1948).

If many of the skills acquired in school are to transfer to new learning or to the solution of new problems, then clearly these skills must be available when needed. But even if a student's learning is adequate at its completion, it need not be adequate, even available, when required (e.g., see Brownell, 1948). Hence, Retention was selected as a fourth criterion.

All the preceding criteria may be classified as "cognitive" learning criteria. Many teachers and educational researchers (e.g., Brown, 1971; Krathwohl, Bloom and Masia, 1964; Messick, 1970) would assert, however, that learning is a phenomena requiring both cognitive and affective criteria to capture its complexity. Accordingly, the following affective criteria were also chosen: Interest in and Attitude toward the algebra both at and two weeks after the completion of its learning. These criteria were selected for the following reasons. First, unlike many affective traits (e.g., values) interests and attitudes might be developed in the brief period over which the experiment was to take place. Second, unlike most affective traits, interests and attitudes can be measured in at least some crude ways (Shaw and Wright, 1967).

Following Getzels (1969), an interest was conceived as a characteristic dispostion of the individual organized through experience which induces the individual to actively seek out particular activities, skills and understandings associated with the object of affect. In the case of this study, interest was defined in terms of the individual's willingness to learn more about the experimental subject and to participate in a number of subject-related activities. An attitude, on the other hand, was conceived as an emotional tendency, organized through experience to act in a characteristic positive or negative way toward the object of affect. In Getzels' scheme the formation of an interest is assumed to be pre-requisite for the formation of an attitude toward a topic or subject.

Instruments: Three parallel forms of twenty-item formative criterion-referenced evaluation instruments were prepared for each unit following procedures outlined by Airasian (1969), Bloom, Hastings and Madaus (1971), and Bormuth (1970). These instruments would be used to determine each student's unit performance level.

Instruments were also developed for each of the learning criterion. Two forms of a twenty-item summative evaluation instrument (Bloom, Hastings, and Madaus, 1971) were developed to test achievement and retention respectively. A ten-item transfer test was devised to test the student's ability to apply some of the major laws of matrix algebra - e.g., the commutative ($A+B = B+A$) and the associative $A+ (B+C) = (A+B) +C$. Finally, Likert-type scales were developed to measure student interest in and attitude toward the algebra. Interest was measured by a scale designed to elicit the student's desire to learn more about various facets of the algebra and to participate in certain activities involving matrix algebra. The attitude scale was adapted from the International Study of Educational Achievement in Mathematics (Husen, 1967) "Attitude toward Mathematics" subscale.[1]

Experimental Procedures: The experiment was performed over one school week under actual school conditions: four sessions of 80 minutes for all students and one session of 40 minutes for students who needed more learning and testing time. At the first session, pretests of achievement, transfer, interest and attitude were administered followed by Unit I. Units II and III were given in the second and third sessions respectively. Post-tests of achievement, transfer, interest and attitude were administered beginning with session four. Two weeks after session four, the retention measure was given and the interest and attitude instruments were re-administered.

Within each of four classes, students were assigned to one of five treatments. Sixteen students were assigned to four experimental treatments where each treatment helped Ss to learn to a different proportion - 65, 75, 85, 95 per cent - of the material in each unit before proceeding to the next. The remaining students were assigned to a control treatment wherein Ss were not required to maintain any particular per unit performance level.

The control and experimental treatments for each unit can be schematized as follows:

Unit Test	Unit Formative Test	Self-directed Review	Parallel Review Test-items	Tutoring	Parallel Review Test-items
← Required →		← As Necessary →			
Control & Experimental		Experimental Only			

In the control treatment, Ss completed the unit programmed textbook, completed the unit test and then, regardless of their score on the unit test, worked on specially assigned homework. In the experimental treatment, Ss completed the unit text and test and then, depending on their score on the test, either moved to the homework or reviewed portions of the unit. If the S had attained his required performance level as indicated by his score on the unit test, he worked on the homework. If not, he reviewed just enough of the unlearned material to bring his performance to standard. Special programmed review materials and individual review prescriptions keyed to these materials were provided. The student could review as much or as little as he

[1] The reliability indices for each learning criterion instrument were:

Achievement	= .84	(Kuder Richardson Formula 21)
Retention	= .81	(" " " ")
Transfer	= .89	(" " " ")
Interest	= .92	(Odd-Even, Split-Half)
Attitude	= .89	(" " " ")

These coefficients are based on a sample size of n = 25.

felt necessary. Upon completion of his review, he was then retested over the reviewed material with new items drawn from the second parallel form of the unit test. If he answered all these items correctly, he was allowed to work on the special homework. If not, he was tutored over the material still unlearned and then retested over this material with items drawn from a third parallel form of the unit test. Pilot testing had shown that this review/correction process would guarantee that virtually all experimental Ss could be helped to reach their required performance level.

Data Gathered: In addition to the pretest, post-test and retention data, the following information was gathered. First, each student's unit performance level before any review/correction. Second, each student's unit performance level after review/correction, if any. Third, the time spent per unit by each student in learning via the textbook and any self-directed review and tutoring. And fourth, each student's interest in and attitude toward the algebra at the completion of each unit.

Data Analysis: Across the four classes, a total of 16 students were assigned to each of the four experimental treatments and 27 students to the control group. However, five experimental and two control students who began the experiment failed to complete it. Eight other experimental students completed the experiment, but were dropped for purposes of data analysis because they consistently exceeded their required performance level (i.e., learned 10 per cent or more material than required) or they consistently failed to attain it (i.e., learned 10 per cent or more less material than required). Consequently, data were analyzed for only 25 subjects in the control group, 12 in the 65 per cent experimental group, 14 in the 75 and 85 per cent groups each, and 11 in the 95 per cent group.

The data were analyzed as follows. First, the mean-scores yielded by each treatment were plotted to investigate the general nature of the relationship between the performance level maintained and student learning as indexed by each criterion. Second, the scores on each criterion measure were analyzed using one-way univariate analysis of variance procedures (Bock, 1963). The least-square estimated effects generated in these analyses and the estimates' standard errors were then used to compare and contrast the effects of the various treatments on each criterion.

Results

Achievement and Retention

As indicated in Figure 1, there was a linear relationship between

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

Insert Figure 1

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

the per unit performance level maintained over the sequence and means scores on the achievement and retention measures. Only maintenance of the 85 and 95 percent levels, however, yielded scores which were significantly higher[1] (achievement: $t_{95} = 2.93$, $t_{85} = 1.93$; retention: $t_{95} = 3.01$, $t_{85} = 1.94$) than the control group's scores.

Besides suggesting that there was some relationship between the maintenance of particular performance levels and the mean level of student achievement, the data also indicated an interesting relationship between the maintenance of the various levels and the variability in student achievement. Table 1 reports the mean achievement test scores and the variance of these scores for each treatment group.[2] Note that as the per unit performance level Ss were asked to maintain increased, mean achievement test scores

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

Insert Table 1

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

rose and the variance of these scores fell. The 85 and 95 percent treatments not only helped students achieve to significantly higher levels than the control treatment, but it also helped homogenize student performance around these high levels.

Transfer

The mean scores of each treatment group on the transfer test are plotted in Figure 2. Here there was no linear relationship between the

[1] $p < .05$. All hypotheses in the study were tested at the .05 level.

[2] Since the mean achievement scores do not approach the test's ceiling, these variances are not articicially restricted.

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

Insert Figure 2

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

performance level maintained and mean scores due primarily to the very low mean score of 75 percent group. Further, only maintenance of the 95 percent standard yielded significantly higher (transfer: $t = 3.02$, $p < .05$) scores than the control treatment.

No completely satisfactory explanation can be given for the low mean transfer score of the 75 percent group. The score might be an artifact of the small sample size and the small number of items (10) in the transfer measure. It might also be attributable to the relatively negative interest in and attitude toward the algebra which, as will be shown shortly, the 75 percent treatment group exhibited after the sequence's completion. The latter explanation is less tenable than the former, however, because the achievement and transfer measures were given together; yet the mean score of the 75 percent group on the achievement measure was not so adversely affected.

<u>Learning Time and Efficiency</u>

Figure 3 illustrates the average total time spent in learning by the

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

Insert Figure 3

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

various treatment groups. As is clear from this figure there was a curvilinear relationship between the performance level maintained and the average total learning time spent. The ANOVA result indicates that all experimental groups spent significantly more ($p < .05$) learning time than the control group.

But one aspect of these time data warrants further analysis. Note that the 75.85, and 95 percent treatment groups all spent approximately the same total amount of learning time despite the fact that the 95 percent group had to learn more material than the 85 percent group and the latter, more material than the 75 percent group. This situation could have occurred only if the 95 percent goup learned more efficiently, i.e, learned more material in a given time, than the 85 percent group and the 85 percent group learned more efficiently than the 75 percent group.

To explore this "efficiency" hypothesis, the total learning time was broken into the time spent in textbook learning and the time spent in correction/review for each unit. The analysis focused on the relationship between the average amount of material learned using only the unit text and the time spent in that learning. Table 2 partially summarizes this relationship.

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

Insert Table 2

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

Note that by Unit III, students in the 95 percent group were spending much less textbook learning time than the other experimental groups and rougly the same time as the control group. Further, by Unit III they were also learning more material, as evidenced by their average formative test score, than any other experimental group and roughly 40 percent more material than the control group. Taken together, therefore, these findings suggest that maintenance of the 95 percent level eventually helped make these students' learning more efficient than the learning of both the control and the other experimental groups.

Interest and Attitude

Figures 4 and 5 are plots of the mean scores for each treatment group

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

Insert Figures 4 and 5

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

on the interest and attitude measures. Figure 4 presents scores on the measures administered with the achievement and transfer instruments; Figure 5 presents scores on the measures administered with the retention instrument. Hereafter, interest in and attitude toward the algebra measured just after its completion will be called "short-term" interest and attitude respectively. Similarly, interest in and attitude toward the algebra two weeks after its completion will be called "long term" interest and attitude respectively.

If the scores of the 75 percent group are disregarded, then in both Figures 4 and 5 there was a curvilinear relationship between the per unit performance level maintained and mean scores on each criterion. In all cases except short-term interest, the scores increased as a function of the level maintained up to the 85 percent level and then dropped off at the 95 percent level. This pattern is especially apparent in the case of long-term attitude.

The ANOVA analyses yielded the following results. On the short-term interest, short-term attitude and long-term interest criterion, both the 85 or 95 percent treatments yielded significantly greater ($p < .05$) scores (short-term interest: $t_{95} = 1.98$, $t_{85} = 1.84$, short-term attitude: $t_{95} = 2.47$, $t_{85} = 2.83$; long-term interest: $t_{95} = 1.97$, $t_{85} = 2.34$) than the control treatment. But the difference in the effects of the 85 and 95 percent treatments was statistically insignificant. On the long-term attitude criterion, however, only the 85 percent treatment yielded significantly greater ($p < .05$) scores (long-term attitude: $t_{85} = 3.42$) than the control treatment.

Discussion

While all these findings must be interpreted cautiously until replicated with a larger sample on a longer learning sequence,[1] on the whole they do suggest that the standard setting approach proposed here is feasible. The maintenance of particular performance levels throughout the instruction did influence the students' future learning as characterized by the selected learning criteria. Further, the maintenance of different levels had different effects on the learning. In particular, maintenance of the 95 percent level best maximized the learning represented by the cognitive criteria while maintenance of the 85 percent level best maximized the learning represented by the affective criteria. Given a model for relating scores on the cognitive criteria to scores on the affective criteria, therefore, it would have been possible to set a mastery standard for the algebra sequence.

Summary

When the task of evaluating student learning is carefully considered, two major problems emerge. One is the gathering of the most appropriate and precise evidence possible about the learning. The other is the setting of performance standards against which this evidence may be weighed and the adequacy of each student's learning judged. This paper has focused on

[1] A replication study is currently under way.

but one facet of the latter problem, viz., the problem of setting performance standards for use in strategies for mastery learning.

The paper began with the argument that a key variable in the design of these strategies are the mastery performance standards which students are helped to attain throughout their instruction. Unless standards can be set whose maintenance does produce the desired end-of-instruction learning outcomes, then these strategies will be misdesigned.

It was then pointed out that presently there are no procedures for setting such standards. Two reasons were given for this situation. First, mastery performance standards must be absolute rather than relative standards for use in interpreting scores on criterion-referenced rather than norm-referenced testing instruments. And second, mastery standards must be set using rational as opposed to arbitrary or irrational techniques. These techniques should yield standards whose superiority as indices of the adequacy of a student's learning can be defended both logically and empirically.

Having established the need for rational procedures for setting mastery standards, next an attempt was made to formulate one such procedure. The approach developed utilizes students' future learning, i.e., their scores on a set of desired, end-of-instruction learning outcomes, as a criterion for determining the mastery performance level which students must attain at any stage in their instruction. To apply the approach, the following steps must be taken. First, the future learning of interest must be specified in terms of a set of defensible learning criteria. Next, a model for interrelating, weighting and combining scores over the various criteria must be developed. This model should capture the wholeness and complexity of the future learning of interest. And finally some statistical technique must be selected for estimating future learning, i.e., scores on the model, as a function of the various performance levels to which the segment or segments of instruction over which the standard will be set might be learned. That performance level whose attainment will yield the greatest estimated future learning is then selected as one's mastery performance standard.

While this approach is neither as simple or as quick as its predecessors, it does have some powerful advantages. First, it introduces a heretofore missing element of objectivity into the mastery standard setting process. It forces one to be explicit about the decision processes by which he arrives at his standards and, thereby, opens the standard setting process up to public scrutiny, challenge, and replication. Second, the approach yields standards whose attainment has clear meaning in terms of the students' future learning. And third, it enables one to optimally design his instruction by establishing a clear relationship between the three major variables which condition how the instruction should proceed: measures of student performance, the learning criteria that are chosen to be optimized, and the available instructional alternatives.

Finally, the paper reported an experiment designed to explore the feasibility of the approach proposed. Perhaps the most basic assumption which underlies this approach is that the performance standard which a student attains over each segment of his instruction has important implications for his realization of the desired, end-of-instruction learning outcomes. The experiment reported was designed to test this assumption.

Ninety-one eighth graders were taught a three-unit sequence in elementary matrix algebra over one school week. The students had been randomly assigned to five treatment groups. The control group learned the algebra under no requirement that they maintain any particular per unit performance level while the experimental groups learned under the requirement that they each maintain a different per unit level. The effects of the control and the experimental treatments on selected, end-of-instruction cognitive and affective learning criteria were then examined. The cognitive criteria were achievement, retention, transfer and learning rate; the affective criteria were interest in attitude toward the algebra at and two weeks after the instruction's termination.

In general, the experiment's results confirmed the assumption tested. The maintenance of particular performance levels throughout the instructional sequence did have significant effects on student learning as characterized by the various learning criteria. Further, the maintenance of particular levels had different effects on different classes of criteria. In particular, the maintenance of one level best maximized scores on the cognitive criteria while the maintenance of another best maximized scores on the affective criteria.

BIBLIOGRAPHY

Airasian, Peter W. Formative Evaluation Instruments: A Construction and Validation of Tests to Evaluate Learning over Short Time Periods. Unpublished Ph. D. Dissertation, University of Chicago, 1969.

Angoff, W. H. "Scales, Norms and Equivalent Scores," in Educational Measurement, Second Edition. Edited by R. L. Thorndike. Washington, D. C. American Council on Education, 1971.

Bhushan, Vidya. Introduction to Matrix Algebra. Honolulu: University of Hawaii, Education Research and Development Center, 1968.

Block, James H. (ed.). Mastery Learning: Theory and Practice. New York: Holt, Rinehart, and Winston, Inc., 1971.

Bloom, Benjamin S. "Learning For Mastery," Evaluation Comment, Vol. 1, No. 2 (1968).

Bloom, Benjamin S. et al., (eds.). Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: David McKay Co., Inc., 1956.

Bloom, B. S., Hastings, T. M. and Madaus, G. F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Bock, R. D. "Programming Univariate and Multivariate Analysis of Variance," Technometrics, 5 (1963), 95-117.

Bormuth, John R. On the Theory of Achievement Test Items. Chicago: University of Chicago Press, 1970.

__________. Development of Standards of Readability: Toward a Rational Criterion of Passage Performance. Final Report, USDHEW, Project No. 9-0237. Chicago: The University of Chicago, 1971.

Brown, George I. Human Teaching for Human Learning. New York: Viking Press, 1971.

Brownell, W. A. "Criteria of Learning in Educational Research," Journal of Educational Psychology, 39 (1948), 170-182.

Cahen, Leonard. "Comments on Professor Messick's Paper," in The Evaluation of Instruction: Issues and Problems. Edited by M. C. Wittrock and David E. Wiley. Holt, Rinehart and Winston, 1970.

Carroll, John B. "A Model of School Learning," Teachers College Record, 64 (1963), 723-733.

Cronbach, L. J. and Snow, R. E. Individual Differences in Learning Ability as a Function of Instructional Variables. Final Report, USOE, Contract No. OEC 4-6-061269-1217. Stanford, California: Stanford University, School of Education, 1969.

Gagné, Robert M. The Conditions of Learning. New York: Holt, Rinehart, and Winston, Inc., 1965.

Gagné, Robert M. "Instructional Variables and Learning Outcomes," in The Evaluation of Instruction: Issues and Problems. Edited by M. C. Wittrock and David E. Wiley. New York: Holt, Rinehart and Winston, 1970.

Getzels, J. W. "A Social Psychology of Education," in The Handbook of Social Psychology, Second Edition, Vol. 5. Edited by Gardner Lindzey and Elliot Aronson, Reading, Mass.: Addison-Wesley Publishing Company, 1969.

Glaser, Robert and Nitko, Anthony J. "Measurement in Learning and Instruction," in Educational Measurement, Second Edition. Edited by Robert L. Thorndike. Washington, D. C.: American Council on Education, 1971.

Husén, Torsten (ed.). International Study of Achievement in Mathematics. Volumes 1 and 2, New York: John Wiley and Sons, 1967.

Krathwohl, D. R., Bloom, B. S., Masia, B. B. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Company, Inc., 1964.

Messick, Samuel. "The Criterion Problem in Evaluation of Instruction: Assessing Possible, Not Just Probable, Intended Outcomes," in The Evaluation of Instruction: Issues and Problems. Edited by M. C. Wittrock and David E. Wiley. New York: Holt, Rinehart and Winston, 1970.

Nedelsky, Leo. "Absolute Grading Standards for Objective Tests," Educational and Psychological Measurement, 14 (1954), 3-19.

Schwab, Joseph J. "The Practical: A Language for Curriculum," School Review, 78 (1969), 1-24.

Scriven, Michael. "The Methodology of Evaluation," in Perspectives of Curriculum Evaluation. Edited by Robert E. Stake. Chicago: Rand McNally, 1967.

Shaw, Marvin E. and Wright, Jack M. Scales for the Measurement of Attitudes. New York: McGraw-Hill, 1967.

Sherman, J. G. "Application of Reinforcement Principles to a College Course." Paper presented at the annual meeting of the American Educational Research Association, New York, New York, 1967.

TABLE 1

THE AVERAGE ACHIEVEMENT TEST SCORES AND THE ACHIEVEMENT TEST SCORE VARIANCES FOR THE CONTROL AND EXPERIMENTAL GROUPS

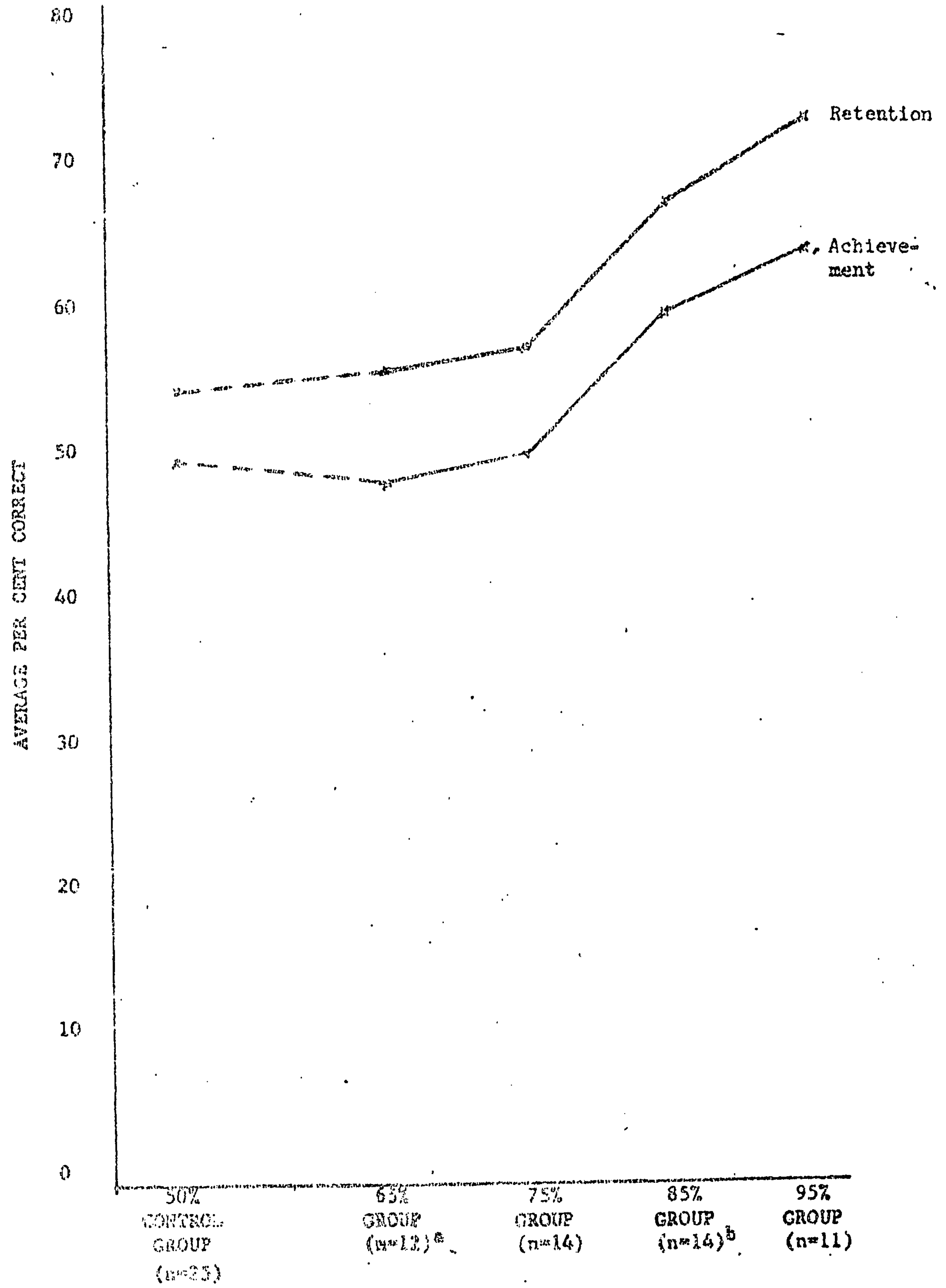
GROUP	AVERAGE ACHIEVEMENT TEST SCORES (PERCENT CORRECT)	VARIANCE OF ACHIEVEMENT TEST SCORES s^2
<u>Experimental</u>		
95 percent (n=11)	64.9	82.8
85 percent (n=14)	60.7	110.2
75 percent (n=14)	50.8	139.2
65 percent (n=12)	49.0	240.2
<u>Control</u> (n=25)	50.5	501.8

TABLE 2

THE AVERAGE AMOUNT OF TIME SPENT IN TEXTBOOK LEARNING PER ALGEBRA UNIT AND THE AVERAGE FORMATIVE TEST SCORES ON UNIT III BEFORE FEEDBACK/CORRECTION FOR THE CONTROL AND EXPERIMENTAL GROUPS

Group	AVERAGE TEXTBOOK LEARNING TIME (in Minutes) Unit I	Unit II	Unit III	AVERAGE FORMATIVE TEST SCORE UNIT VI (Mean Percent correct)
Experimental				
95 percent (n=11)	11.4	14.2	25.8	74.4
85 percent (n=14	11.4	15.0	29.4	63.4
75 percent (n=14)	11.2	15.2	29.1	56.5
65 percent (n=12)	11.1	14.3	27.1	63.7
Control (n=25)	11.1	12.7	25.3	54.2

Fig. 1 ---Average achievement and retention scores for the control and experimental groups.



[a] The n was 11 for the retention mean score.

[b] The n was 12 for the retention mean score.

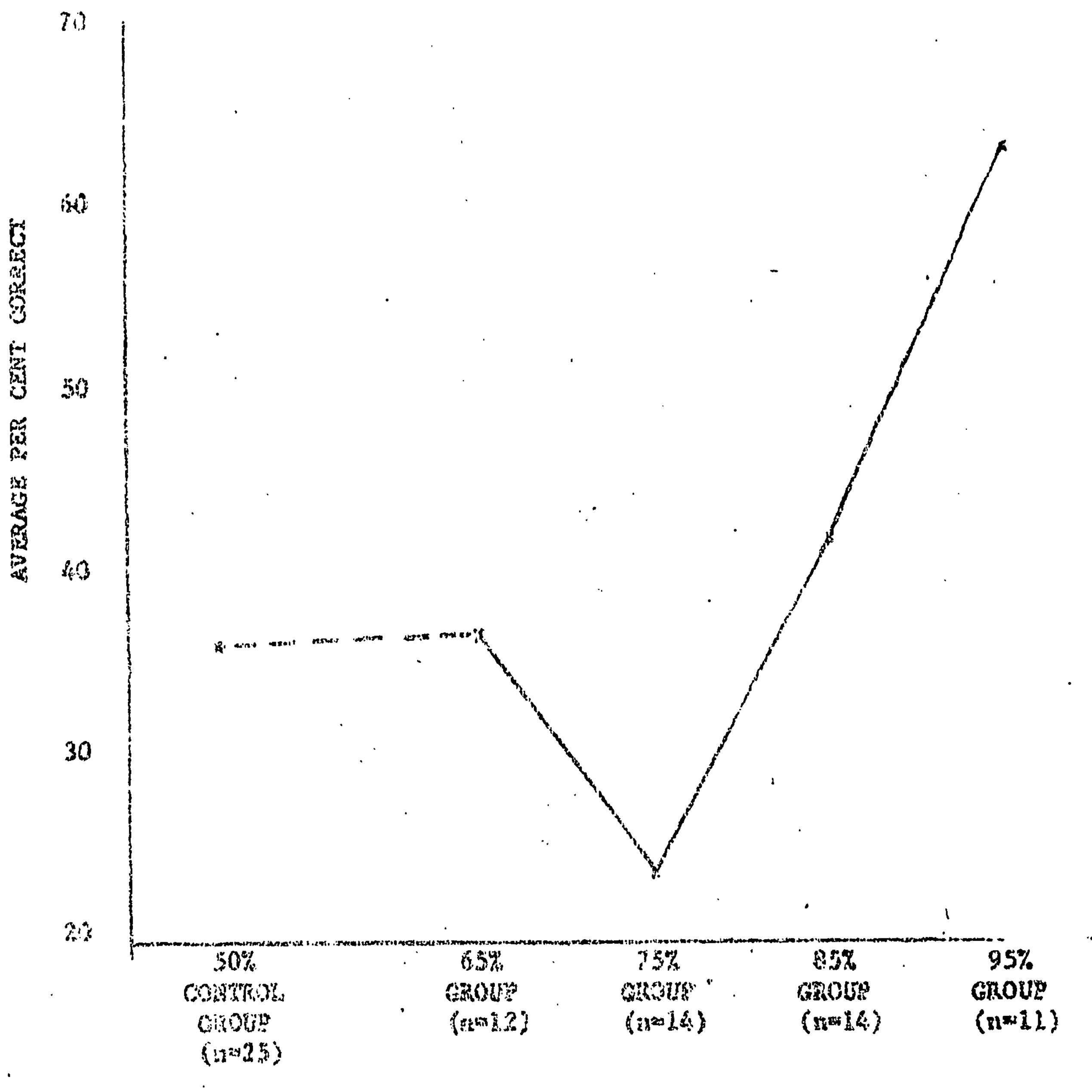


Fig. 1.--Average transfer scores for the control and Experimental groups.

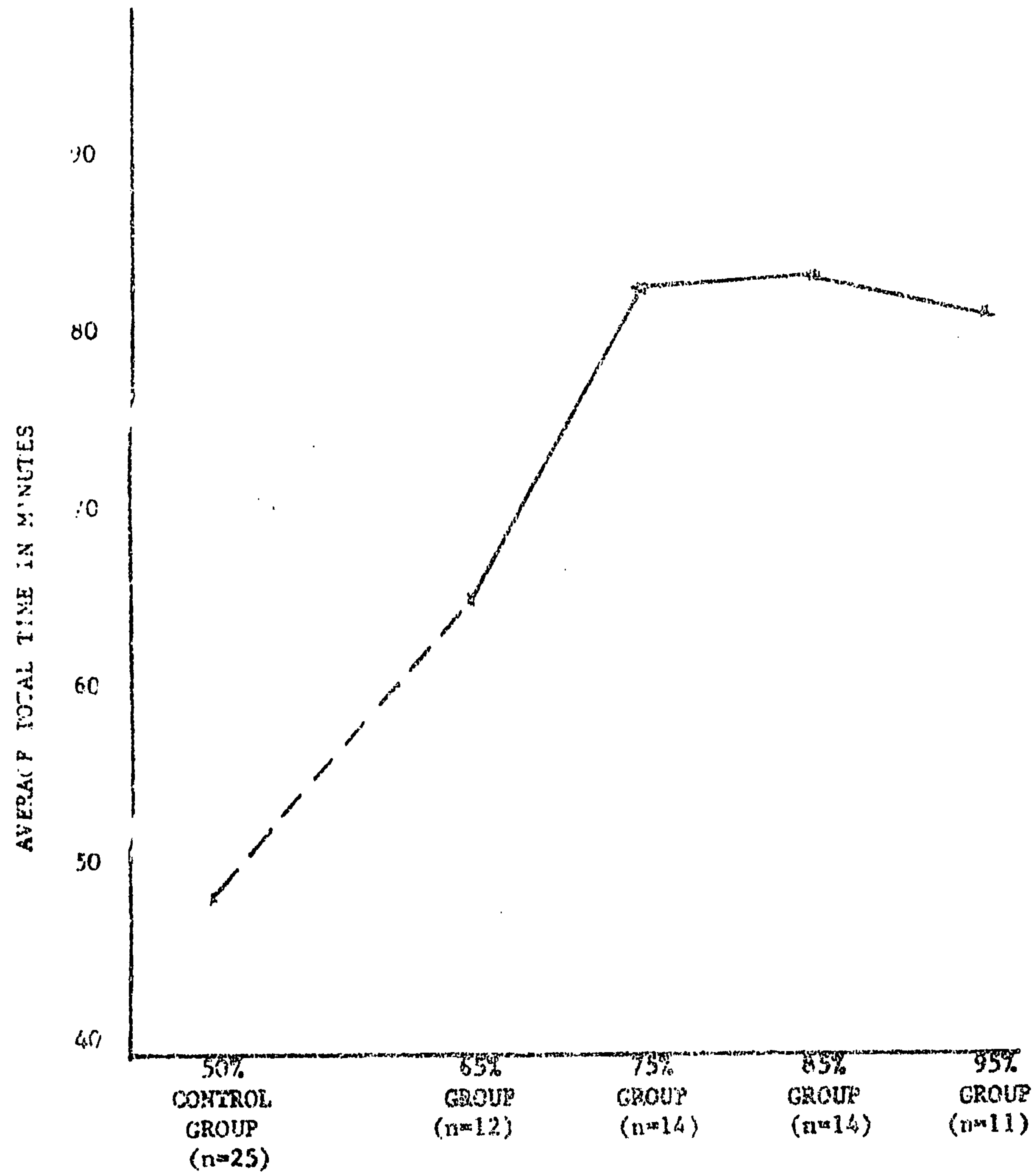


Fig. 3.--Average Total Learning Time for the Control and Experimental Groups.

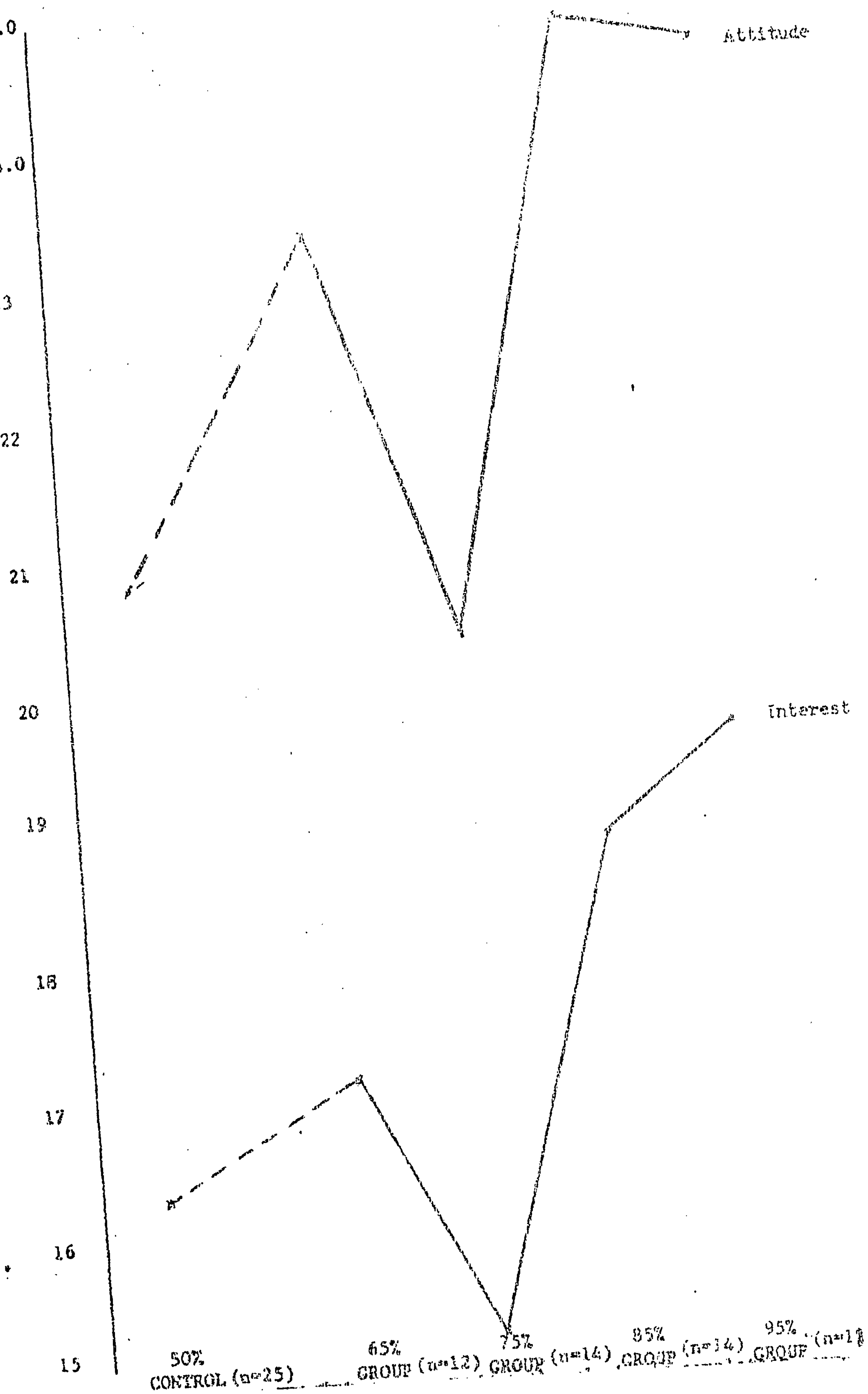


Fig. 4 — Average short term interest and attitude scores for the control and experimental groups.

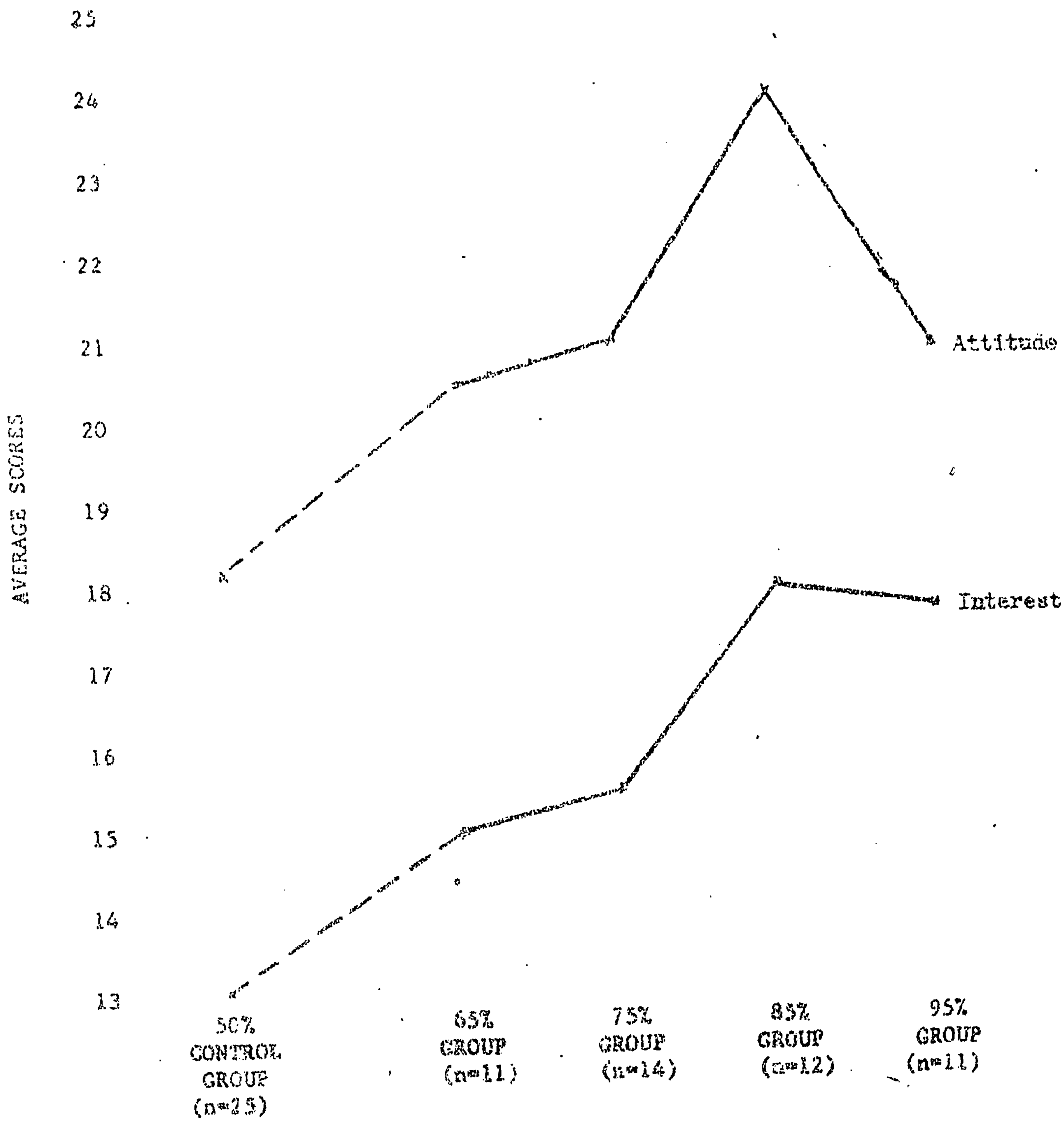


Fig. 5 --- Average long-term interest and attitude scores for the control and experimental groups.