DOCUMENT RESUME

ED 108 074 95 CG 009 893

AUTHOR McMullen, David W.
TITLE Realistic Self-Assessment of Knowledge and Competence. Final Report.
INSTITUTION State Univ. of New York, Stony Brook.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Regional Research Program.
BUREAU NO BR-2-B-071
PUB DATE Dec 73
GRANT OEG-2-72-2B071
NOTE 53p.; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.76 PLUS POSTAGE. HC Not Available from EDRS.
DESCRIPTORS Behavior; *Computer Oriented Programs; Feedback; *Junior High School Students; *Knowledge Level; Probability; Research Projects; *Self Evaluation; Training Objectives; *Training Techniques

ABSTRACT

Decision theoretic testing was used to explore whether junior high students could improve their tendencies to make realistic assessments of what they knew. In nine sessions over a period of three weeks, 49 seventh graders used computer terminals to record probability values in the form of logarithmic equivalents for each alternative in a randomized set of multiple-choice math problems. Students were assigned to three treatment groups. One worked individually, one worked individually with feedback from the computer, and the third group worked as teams with feedback from the computer. Analysis of the training sessions and posttests revealed differences in the stability of assessment behaviors but a common tendency to be overconfident, to be less realistic as problems become more difficult, and to distort the value of one's knowledge when working as teams. Realism training was most effective when explicit feedback was provided and the achievement level was low. (Author/PC)

FINAL REPORT

Project No. 2B071
Grant No. OEG-2-72-2B071

David W. McMullen
State University of New York at Stony Brook
Stony Brook, New York 11790

REALISTIC SELF-ASSESSMENT OF KNOWLEDGE AND COMPETENCE

December, 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education

National Center for Educational Research and Development
(Regional Research Program)

ABSTRACT

Decision theoretic testing is used to explore whether junior high students can improve their tendencies to make realistic assessments of what they know, i.e. to assign subjective probabilities congruent with their information and reasoning about the problem at hand. In nine sessions over a period of three weeks, 49 seventh graders used computer terminals to record probability values in the form of logarithmic equivalents for each alternative in a randomized set of multiple-choice math problems. Ss viewed problems, calculated values on a set of panels called a SCoRule, entered a value for each alternative, and received immediate feedback after each problem.

Students were assigned to three treatment groups. One worked individually, one worked individually with explicit feedback about observed realism in the form of a computer-generated graph at the end of each session, and the third group worked as teams of size three or four, each team sharing a terminal and receiving the graph.

A paper-and-pencil posttest showed a slight gain in realism over an equivalent pretest, but the difference was not statistically significant. A dramatic shift toward defined realism (observed realism within five degrees of perfect realism) did occur, however, among low achievers in the second treatment group. A second posttest, administered as teams of size three instead of individually, contained harder items than the first posttest, and realism loss was significant, especially for the team study treatment group.

Analysis of the training sessions and posttests revealed differences in the stability of assessment behaviors but a common tendency to be overconfident, to be less realistic as problems become more difficult, and to distort the value of one's knowledge when working as teams. Realism training was most effective when explicit feedback (i.e. the graph) was provided and the achievement level was low. Time constraints, however, may have reduced effectiveness, and an extended program of realism training may be needed in order to modify long-established tendencies.

Final Report

Project No. 2B071
Grant No. OEG-2-72-2B071

Realistic Self-Assessment of Knowledge and Competence

David W. McMullen

State University of New York
at Stony Brook

Stony Brook, New York

December 1, 1973

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

ACKNOWLEDGEMENTS

The project was possible only through the combined efforts of a number of individuals and the cooperation of the Instructional Resources Center at Stony Brook. The author is especially indebted to Dr. Emir Shuford and Dr. Robert Brennan for the assistance they have provided throughout the project. Dr. Shuford was instrumental in initiating and guiding the project. Dr. Brennan made numerous contributions in both the analysis and interpretation of data as well as in suggesting procedures.

Betty Weneser, Paul Siegel and Kathy Lum of the Stony Brook Computer-Assisted Instruction Laboratory directed and produced the programming and made many helpful suggestions. James Rubin assisted in the data analysis. To these and to Dr. Ed Lambe, Director of the Instructional Resources Center, I extend my thanks.

I am also grateful for the cooperation of the Three Village School District on Long Island, and especially William Braun and Kennth Swinson for the participation of their classes.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

I. Statement of the Problem

Educators frequently assert that along with knowing facts, principles, and skills it is important that students know what they don't know. A student who is unaware of the limitations of his knowledge is unprepared to make sound judgments in situations of uncertainty. By placing either too much or too little value on what he knows, he will make decisions that fail to take full advantage of his state of knowledge and may even lead to serious losses.

> Above all, the educational advantage of training people - possibly beginning in early childhood - to assay the strengths of their own opinions and to meet risk with judgment seems inestimable. The usual tests and the language habits of our culture tend to promote confusion between certainty and belief. They encourage both the vice of acting and speaking as though we were certain when we are only fairly sure and that of acting and speaking as though the opinions we do have were worthless when they are not very strong. (Savage, p. 800)

One type of the "usual test" that may foster confusion between belief and certainty is the simple choice method in which a single alternative receives a probability of 1.00 (certainty) even when the alternative is given only a slight preference over the others. By excluding the other alternatives in the presence of uncertainty, the student learns to assign both more worth and less worth to what he knows than he should. He passes over the value of information that favors excluded alternatives and he gives disproportionate value to information that supports his choice.

Applications of decision theory to testing (Shuford, Albert & Massengill, 1966) havd led to the development of new methods of responding in which students distribute probabilities over the range of alternatives. Besides improving predictive validity and increasing the amount of information available to the teacher, decision-theoretic tests generally require fewer items than conventional tests. More important to this study, however, is the capability provided by decision-theoretic tests for measuring student realism, a new dimension in student achievement. Realism is a congruence between the probability one assigns to an event and the probability that the event will occur given what one knows.

A student with a tendency to make realistic assessments of his state of knowledge and competence knows precisely the worth of his information and reasoning in making decisions and solving problems. As a result, he recognizes his level of mastery of a topic. He is aware of the worth of his judgment and aware of his lack of critical information. He knows when his reasoning is logically tight and when it is not.

Measuring realism is analogous to assessing the accuracy of weather forecasting. Probabilistic predictions of rain have become common in weather forecasts. If rain falls with approximately the frequency estimated by the forecaster, he uses his information realistically. Thus, predictions of an 80% chance of rain (p=.80) can be grouped into days when rain fell and days when no rain fell. If the proportion of days when rain fell to the total of the two groups approximates .80, a data point can be plotted on a 45 degree line indicating perfect realism (Fig. 1). If, however, predictions of a 50% chance of rain are associated with a set of days in which the proportion of rain days was only .30, a point would appear below perfect realism. With enough data points, it is possible to plot probability assignments against relative frequency of rain and arrive at a least-squares estimate of the forecaster's realism line. Similarly, a student deviates from perfect realism to the degree that his reported probabilities fail to match his true probabilities as reflected by the relative frequency with which his knowledge allows him to ascertain the outcome of an event (e.g. the correct solution).

The relationship between realism and achievement is depicted by the <u>information square</u> (Fig. 2), in which the horizontal dimension is realism and the vertical dimension is achievement. When the student's actual state of knowledge lies below the level at which he estimates or perceives that knowledge, the student is overconfident, in the sense that he overvalues what he knows. When perceived knowledge is below actual knowledge, he is underconfident. To be realistic the student must exhibit a tendency to evaluate his knowledge at the level of actual knowledge. Note that perfect realism is possible at any level of achievement. As an interpretation of the information square, Shuford (1973) cites this Arabian proverb:

> He who knows and knows that he knows,
> He is wise, follow him.
>
> He who knows and knows not that he knows,
> He is asleep, awaken him.
>
> He who knows not and knows not that he knows not,
> He is a fool, shun him.
>
> He who knows not and knows that he knows not,
> He is a child, teach him.

As the proverb suggests, one consequence of realism may be more efficient learning since the student knows where to concentrate his efforts. A tendency toward realism may have

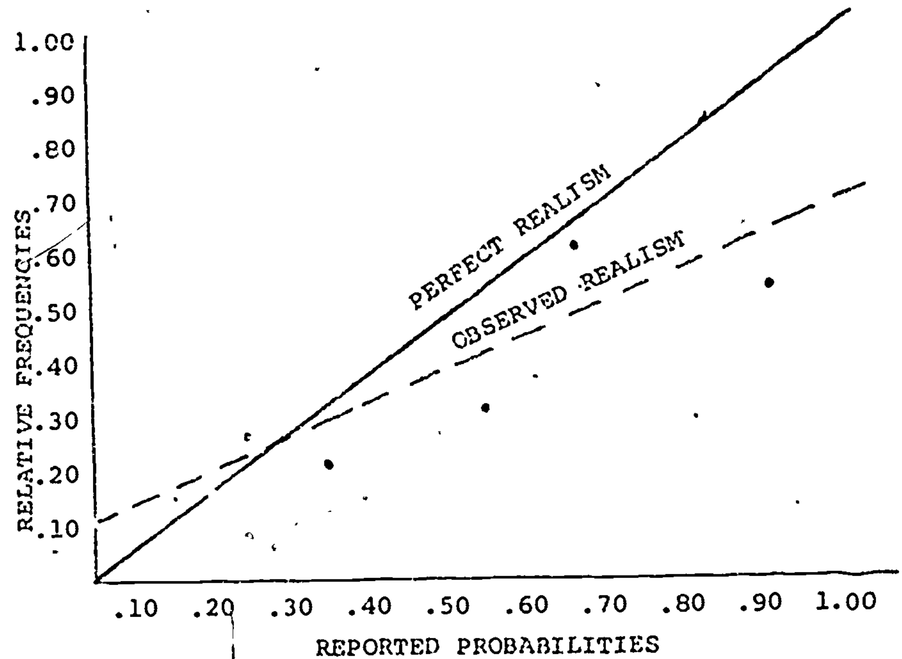

Fig. 1. Realism graph, showing relationships between reported probabilities and relative frequencies (state probabilities).

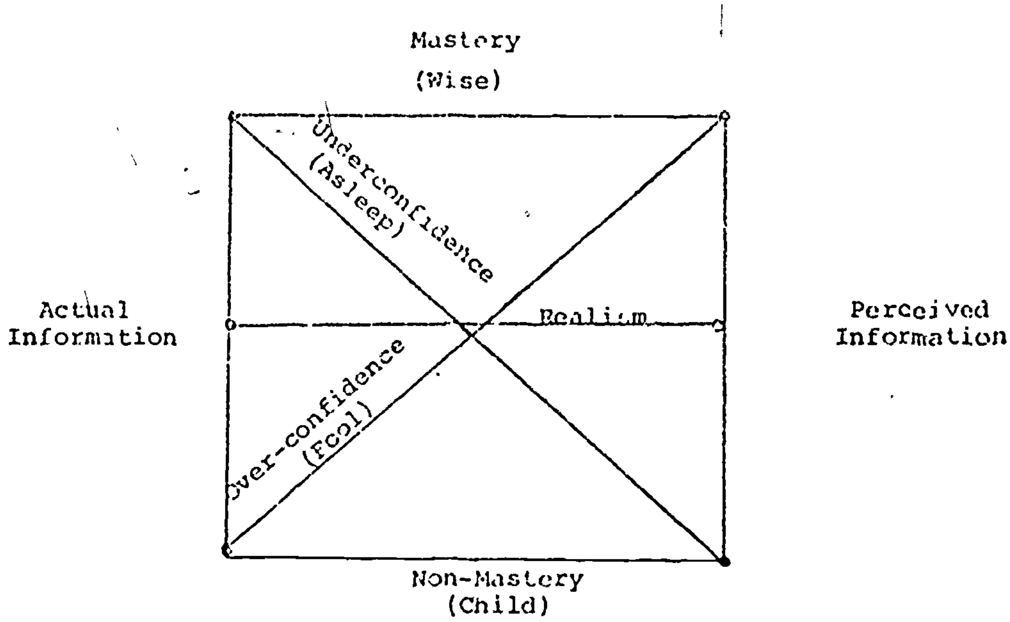

Fig. 2. Information square relating realism and achievement (see text for explanation)

a number of other important benefits, both cognitive and affective. The realistic student has learned to make critical evaluations of knowledge claims. He has learned to approach problems without bias. He has made decisions that capitalize on his available resources, neither exaggerating nor depreciating them. Under conditions of risk and uncertainty, he manages to keep what he knows in perspective. As preparation for living in the uncertain context outside of school, realism training may have far-reaching significance.

The objective of this study was to examine the extent to which realism training could be effective with a group of seventh grade math students. Realism training in this instance was a series of short but intensive sessions in mathematical problem-solving, using the computer to administer and monitor decision-theoretic testing and to provide immediate feedback to the students. An effective training program is one that results in a significant improvement in realism for many of the students. According to the little data available, the degree of measured realism for junior high students appears to be relatively stable over time. These findings, however, came from unsystematic and non-statistical studies. It was the intent of this research to investigate a concentrated effort to improve students' assessment tendencies.

II. Procedure

The experiment was based on previous research with the "SCoRule," a set of panels by which to translate probability assessments into logarithmic equivalents (Shuford, 1970). For example, on a four-alternative item with a single correct answer, each alternative has a .25 probability of being correct if each alternative is equally likely. The logarithmic equivalent on the SCoRule, however, is 70. An alternative with only .10 probability nevertheless has a SCoRule value of 50, while the value for .50 probability is 85 and .60 probability is equivalent to 89. In other words, the logarithmic transformation returns proportionately large SCoRule values for lower probability values but SCoRule values increase only slightly in the upper range of probability values.

The student assigns SCoRule values to each alternative and receives the value he has assigned to the alternative that is correct. Because his score is a log value, the student loses little by slightly reducing the probability he attaches to the alternative he prefers, but by slightly increasing lower probability values he stands to make a substantial SCoRule gain if one of those alternatives turns out to be correct. The SCoRule scoring system therefore makes it worthwhile to the student to reveal his uncertainty.

A. Subjects

Two seventh grade classes from a suburban junior high school participated during their regular 45-minute math periods. One class, taught by the department chairman, was considered higher in achievement than the other class. A total of 54 subjects participated in the experiment but because of absences and unexpected circumstances only 49 provided adequate data for analysis, 26 in the high-achieving class and 23 in the low-achieving class.

B. Materials

The experiment used six forms of the mathematical problem-solving sections of the Iowa Test of Basic Skills. Forms 1-4 were pooled and then divided into 15 levels of difficulty on the basis of item analysis information provided with the test. The first item at each level was assigned to the problem set for the first training session, the second item at each level to the problem set for the second session, etc., until nine problem sets were prepared for nine training sessions. Form 5 was administered as the pretest and Form 6 as the first posttest. A second posttest consisted of the last 21 items in Form 5, which no S had reached during the pretest.

At the time of the pretest a SCoRule Procedures Test was administered in order to measure ability to use the SCoRule after pretraining (Appendix A).

C. Experimental Task

Nine training sessions were conducted in the Computer-Assisted Instruction Laboratory at Stony Brook during a three-week period. At each session S used a scope-and-keyboard computer terminal to register SCoRule values opposite each of four alternatives for as many items as time permitted. The computer monitored the input, converting the SCoRule log values to probabilities. If the probabilities assigned to an item did not sum to 1.00 within a tolerance of .04, the computer required the student to retype his SCoRule values.

Because of time required for busing to and from the lab, each session was 20-25 minutes in length. At the end of each session the computer generated "realism graphs" like Fig. 1 with two exceptions. First, the vertical axis was changed so that an observed realism line deviating from perfect realism in the direction of overconfidence appeared with the upper part of the line above the perfect realism line, while underconfidence was graphed below perfect realism. This change made the graph easier to interpret to Ss since they could readily see if they were overvaluing or undervaluing their knowledge. Second, additional feedback appeared on the graph: (1) one of the following words: "realistic," "overconfident," "underconfident;" (2) the number of problems completed; and (3) the accumulated SCoRule score for the session. "Realistic" appeared if the deviation between observed realism and perfect realism (cf. Fig. 1) was five degrees or less in either direction. The decision to set five degrees as the tolerance range was the result of a series of simulations on a PDP-10 analyzing a variety of SCoRule settings. The simulation procedure, briefly, was to construct several 15x4 matrices, each representing 15 four-alternative items and log (SCoRule) values that yielded an observed realism line with a slope of 1.00. Slight changes in values were made until the settings indicated to the experimenter a clear case of either overconfidence or underconfidence. Departure from realism consistently appeared when the angle between perfect and observed realism exceeded five degrees, or defined realism.

Though emphasis was on achieving realism, each session provided the cumulative number of items and the cumulative score at both the end of the session and after each item. The purpose was to encourage S to work rapidly on a number of items rather than lingering over a single problem until extensive calculations removed all uncertainty. S was told to expect to be somewhat uncertain in registering decisions but to do as many items as possible while using what he did know as wisely as he could. The score was designed to discourage carelessness on the part of those who worked rapidly as well as to provide an incentive for attempting to do all 15 items.

Ss were instructed that the preferred strategy in attacking a problem was to set the SCoRule to indicate no knowledge, i.e. each alternative with .25 probability, or SCoRule values of 70-70-70-70. Values could then be raised or lowered depending on S's knowledge with regard to given alternatives. The cumulative score was also designed to discourage S from simply leaving the SCoRule at the no-knowledge level, thereby showing excessive caution.

When a graph appeared at the conclusion of a session, the experimenter attempted to interpret the results with comments such as, "'Overconfident' means that you are placing too many points on what you consider the best alternative," "'Underconfident' means that you need to give more points to the alternative you think is best," "'Realistic' means that you are giving just about the right amount of value or your SCoRule to each alternative." Suggestions were sometimes made to "spread out points" (overconfidence) or to "lump points" (underconfidence).

D. Design

To assess whether the training sessions were associated with an improvement in realism, the study used a pretest-posttest design accompanied by trend data from the sessions themselves. Since the study was exploratory, the primary objective was to determine whether any change would occur, leaving to a later study comparisons with a control group.

The pretest and first posttest were both taken individually with the materials described above. The second posttest was administered to Ss working as teams of three on the day following the first posttest. Though they were encouraged to consult with other team members, they were told to record SCORULE values separately, making their own decisions. As noted earlier, the problems on the second posttest were more difficult than those on the other two tests. Teams were formed by the teachers in order to secure as much participation and cooperation in the groups as possible.

A secondary objective, however, was to examine the relative effects of exposure to the graph and team problem-solving. Three groups were formed in each class. The groups differed in the type of support given during realism training. Group 1 received the least support, Group 2 received explicit feedback (the graph), and Group 3 (team study) both explicit feedback and peer interaction. Whether or not peer interaction was in fact supportive in attaining realism was one question the study sought to examine.

Each teacher assigned one third of his class to the team study group in order to keep each team congenial. Insofar as possible, teachers made each team representative of the class, with high achievers balanced by low achievers. One three-member team in each class was male and one female. An additional female team in the larger class had four members, one of whom enrolled after the original assignments were made but before training began.

The remaining two thirds of the total sample were divided into four blocks before being randomly assigned to work alone either with or without exposure to the graph. The pretest was used to make a preliminary estimate of achievement level by using the number of items completed as an index, since pretest scores could not then be obtained without long and complicated analyses. (Since then a FORTRAN program called DEC-TEST (Brennan, 1973) has been written by Dr. Robert Brennan of Stony Brook by which to make rapid analyses of decision-theoretic tests.) The index correlated .64 with the math applications section of the California Achievement Test, which was available for only 26 of the students, and .[illegible] with the reading score of that test, which was available for 27. (The two parts of the test correlated .[illegible].)

[illegible] familiarity with the SCoRule might also influence the effectiveness of realism training another index was computed using scores [illegible] Test" developed by the experimenter (Appendix A). Correlation between this familiarity index and the number done, the math applications and reading scores were .10, .26, and .24.

The mean number of items completed on the pretest was 20.6. The mean SCoRule test score was 7.0. If S was above the mean in both cases, he or she was assigned to the high-achiever/high-SCoRule familiarity block. Those below the mean were considered low, either in achievement level or SCoRule familiarity, or both. Within the four blocks that were formed, Ss were also blocked by class and by sex. Those in the High class (cells A and C in Figure 3) were randomly assigned to one of the two individual study treatments, with the number of males in each cell kept equal. Similarly low class Ss were randomly assigned to cells B and D. Figure 3 shows cell Ns, sex composition, and mean scores on blocking indices for the full 2x3 design. Analysis of variance disclosed that the treatments did not differ on the achievement index (F=1.26), nor did the cells differ on either CAT math applications scores (F=.56) or reading scores (F=.19). Analysis of the SCoRule familiarity index revealed more variability (F=3.34, $p < .05$), although individual study treatments were similar (F=.31).

III. Results

Pretest-Posttest Comparisons

The major dependent variable was the degree of the angle between observed realism and perfect realism. Observed realism was computed by the procedure described in Appendix B. Table 1 presents the mean angle of deviation for each treatment group and for the total sample on the pretest and the two posttests. The results show essentially no difference between pretest and posttest behavior except for Group 3 on the second posttest. Though the overall mean and the means for Groups 1 and 2 show a slight gain in realism on Posttest 1, the Wilcoxon signed-ranks test for two matched samples failed to disclose significance (z=.98, .11, and 1.17 respectively).

While the magnitude of realism gain on the first posttest is small, the direction of the change in standard deviations is consistently toward improvement. Even the second posttest, which combined student teams with harder problems and showed a sharp loss in realism, indicated that variability decreased following training.

An estimate of the proportion correct can be obtained by considering the most highly valued alternative as the single response S would have made under classical testing

CLASS	INDIVIDUAL STUDY NO GRAPH 1	INDIVIDUAL STUDY GRAPH 2	TEAM STUDY GRAPH 3
HIGH	A 21.4 N=7 M=3 F=4 8.5	C 8.9 N=9 M=3 F=6 7.8	E 23.1 N=10 M=3 F=7 6.6
LOW	B 18.8 N=8 M=3 F=5 6.3	D 19.0 N=9 M=4 F=5 8.1	F 23.0 N=6 M=3 F=3 3.4

Figure 3. Sample size, mean number of items completed on pretest (upper right corner), and mean score on SCoRule Procedures Test (lower right corner) in each cell of the design used to compare treatment effects (M=male, F=female).

Table 1

Mean Angle of Deviation from Perfect Realism on Pretest and Posttests By Treatment Groups in Degrees (Standard Deviation in Parentheses)

Treatment	Pretest	First Posttest	Realism Gain/Loss	Second Posttest	Realism Gain/Loss over Pretest
Group 1 (individual study without graph) N=15	14.9 (11.0)	11.1 (7.2)	+3.8 (+3.8)	17.4 (6.2)	-2.5 (+4.8)
Group 2 (individual study with graph) N=18	14.4 (9.9)	10.8 (9.1)	+3.6 (+ .8)	18.1 (7.7)	-3.7 (+2.2)
Group 3 (team study with graph) N=16	10.6 (12.6)	12.2 (9.4)	-1.6 (+3.2)	22.2 (9.6)	-11.6** (+3.0)
Total Sample N=49	13.3 (11.3)	11.3 (9.7)	+2.0 (+2.6)	19.3 (8.3)	-6.0** (+3.0)

** $p < .01$, Wilcoxen signed-ranks test conditions, i.e. the choice method. In the case of tied values, one is picked at random in order to estimate $\underline{S}$'s response. Table 2 reveals that the pretest and first posttest were of comparable difficulty but the second posttest was considerably harder. Variability decreased on the posttests as it did with respect to the realism angle.

Estimates of proportion correct confirmed the teachers' judgment that the two classes differed in achievement level. The low class remained at least 15 points behind the high class but showed a modest relative gain (Table 3).

What gain in realism that did occur on the first posttest came primarily from the low-achieving class, as shown in Table 4. Although the gain of 4.2 degrees was not statistically significant (z=.70), the overall pattern indicates that the low class tended to respond more favorably to realism training than did the high class. The pattern also appears when the classes are compared according to the proportion in each class whose observed realism line fell within five degrees of the perfect realism line, i.e. the proportion who were "realistic" within the range of tolerance used in the experiment (defined realism). In the high class the proportion was the same for both the pretest and the first posttest, while the proportion in the low class more than doubled (Table 5).

The proportion who attained defined realism increased the most for the group working individually and receiving the realism graph (Table 6). Team study with the graph, however, was associated with a decrease. Substantial posttest gain really occurred in only one cell in the original design: low achievers under individual study with the graph (treatment 2). Table 7 presents the distribution of $\underline{S}$s on the pretest and first posttest in the three categories that were used for feedback with the realism graph. While the number achieving defined realism is stable from pretest to posttest for five of the six cells, underconfident and overconfident students move sharply toward realism in the low-class-graph feedback group.

<u>Training Session Trends</u>

<u>Realism and difficulty</u>. Estimates of realism during the C.A.I. sessions are based on fewer units of analysis because of two major factors: absenteeism and the fact that Group 3 shared terminals. In general, $\underline{S}$s showed less realistic behavior during the nine sessions than on the pretest and the first posttest, but mean fluctuations remained within the range of realism bounded by the means of the two

Table 2

Mean Estimated Proportion Correct
On Pretest and Posttests
By Treatment Groups
(Standard Deviations in Parentheses)

Treatment	Pretest	First Posttest	Second Posttest
Group 1 (Individual study without graph) (N=15)	.62 (.18)	.61 (.13)	.48 (.11)
Group 2 (Individual study with graph) (N=18)	.55 (.21)	.60 (.17)	.52 (.11)
Group 3 (Team study with graph) (N=16)	.66 (.21)	.65 (.16)	.46 (.12)
Total Sample (N=49)	.61 (.20)	.62 (.16)	.49 (.11)

Table 3

Mean Estimated Proportion Correct
on Pretest and Posttests
By Classes
(Standard Deviation in Parentheses)

Class	Pretest	First Posttest	Second Posttest
High (N=26)	.70 (.15)	.69 (.13)	.56 (.08)
Low (N=23)	.50 (.20)	.54 (.15)	.39 (.08)
Total Sample (N=49)	.61 (.20)	.62 (.16)	.49 (.11)

Table 4

Mean Angle of Deviation from Perfect Realism on Pretest and Posttests By Classes in Degrees (Standard Deviation in Parentheses)

Class	Pretest	First Posttest	Realism Gain/Loss	Second Posttest	Realism Gain/Loss over Pretest
High (N=26)	9.2 (7.5)	9.1 (5.9)	+ .1 (+1.6)	16.2 (4.4)	-7.0** (+3.1)
Low (N=23)	18.0 (13.0)	13.8 (10.5)	+4.2 (+2.5)	23.5 (10.3)	-5.5 (+2.7)
Total Sample (N=49)	13.3 (11.3)	11.3 (8.7)	+2.0 (+2.6)	19.3 (8.3)	-6.0** (+3.0)

** $p < .01$, Wilcoxen signed-ranks test

Table 5

Proportion Attaining Defined Realism (Within 5 Degrees of Perfect Realism) By Classes

Class	Pretest	Posttest	Second Posttest
High (N=26)	.35	.35	.00
Low (N=23)	.13	.30	.00
Total Sample (N=49)	.25	.33	.00

Table 6

Proportion Attaining Defined Realism
(Within 5 Degrees of Perfect Realism)
By Treatment Groups

Treatment	Pretest	First Posttest	Second Posttest
Group 1 (individual study without graph) (N=15)	.13	.20	.00
Group 2 (individual study with graph) (N=18	.11	.39	.00
Group 3 (team study with graph) (N=16)	.50	.38	.00
Total Sample (N=49)	.25	.33	.00

Table 7.

Distribution in Three Categories of Observed Realism
By Treatment Groups within Classes
on Pretest and First Posttest

	Treatment Group	N	Pretest Under-conf.	Pretest Realistic*	Pretest Over-conf.	First Posttest Under conf.	First Posttest Realistic*	First Posttest Over conf.
High Class	1	7	1	1	5	0	2	5
	2	9	0	2	7	0	2	7
	3	10	0	6	4	0	5	5
	All	26	1	9	16	0	9	17
Low Class	1	8	1	1	6	2	1	5
	2	9	2	0	7	1	5	3
	3	6	0	2	4	0	1	5
	All	23	3	3	17	3	7	13
Total Sample		49	4	12	33	3	16	30

*observed realism within five degrees of perfect realism posttests (Figure 4). Both the smaller sample sizes and the small number of items per session may be responsible for much of the fluctuations through measurement error (see Appendix C for session data and for the number of items completed during both tests and sessions).

Fluctuations, however, do follow a consistent pattern if they are viewed in connection with the difficulty of the items. The estimated proportion correct across all Ss also fluctuates within the range bounded by the means of the two posttests (Figure 5). Realism and difficulty, however, fluctuate in opposite directions, as shown by Figure 6. Increased difficulty is associated with less realism.

Most of the fluctuation occurred in the low class, which was consistently less realistic than the high class (Figure 7). Though the high class showed a slight tendency toward greater realism, its performance was generally stable. Both classes exhibited variation from session to session in the estimated proportion correct (Figure 8). The low class consistently scored below the high group and performed less stably. Neither class scored especially high; even the high class found the tests and sessions difficult. The strong relationship between realism and difficulty is apparent in Figure 9, which shows that for the low class item difficulty and realism angle were nearly mirror images. When the randomly selected set of training items was relatively difficult, Ss were relatively unrealistic, but realism over a session improved when items were less difficult.

The comparison of the performance of the three treatment groups during training sessions is complicated by the small N in Group 3 that resulted from sharing terminals. Group 1 (individual study without the graph) exhibited the most realistic behavior until midway through the experiment, when Group 3 (team study with the graph) took the lead (Figure 10). These two groups, however, also tended to find the items less difficult (Figure 11). The realism and difficulty curves of Group 2 are another case of a near mirror image (Figure 12).

The effect of difficulty on realism was also examined by comparing easy and hard items that appeared during training. Because of random presentation within the problem set for each session, it was possible to divide each set into comparable subsets: items from the easiest seven levels and items from the hardest seven levels. While in a few instances S did not encounter items from one of the subsets, the number of Ss and the mean number of items actually presented were approximately equal for the two subsets on a given training session (see Appendix D). The mean realism angles, however, were highly divergent for the two subsets (Figure 12). Except for the third session, deviation from realism was consistently greater for harder items than for easier items.

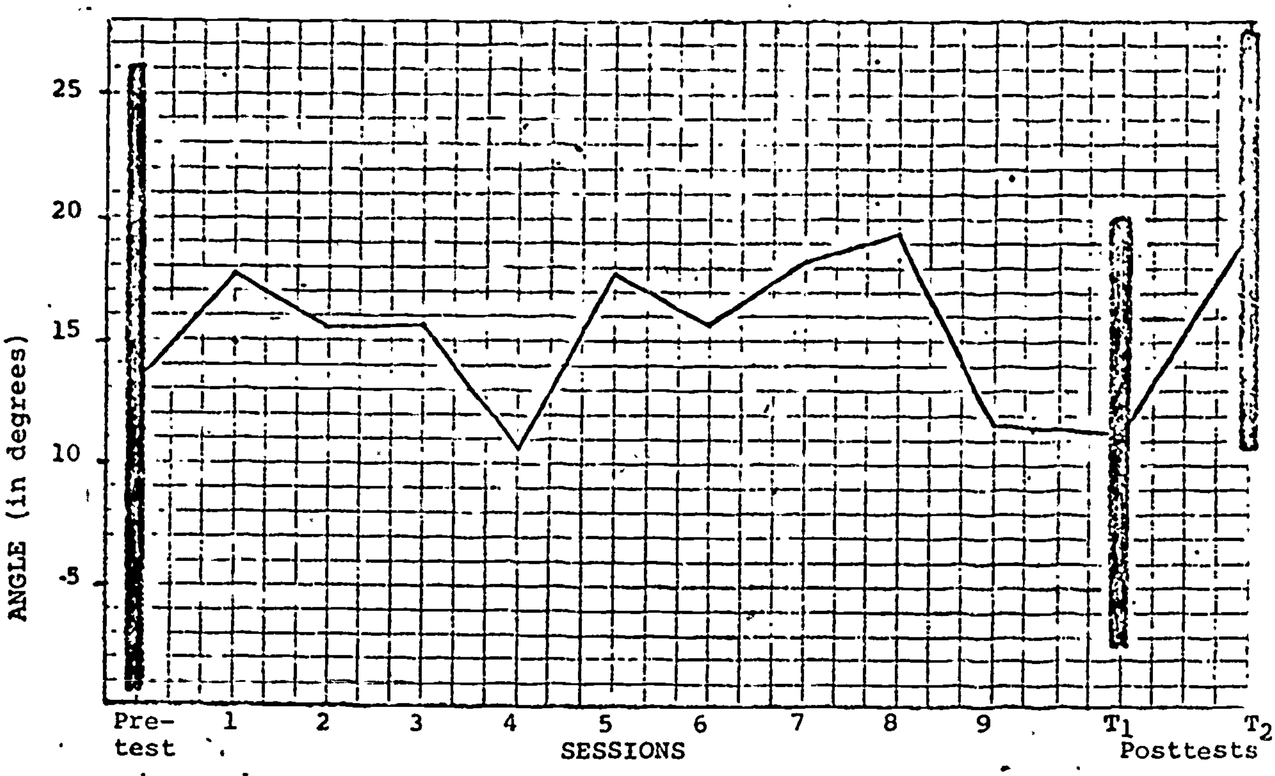

Figure 4. Angle between ideal and observed realism for total sample (see Appendix C for Ns).

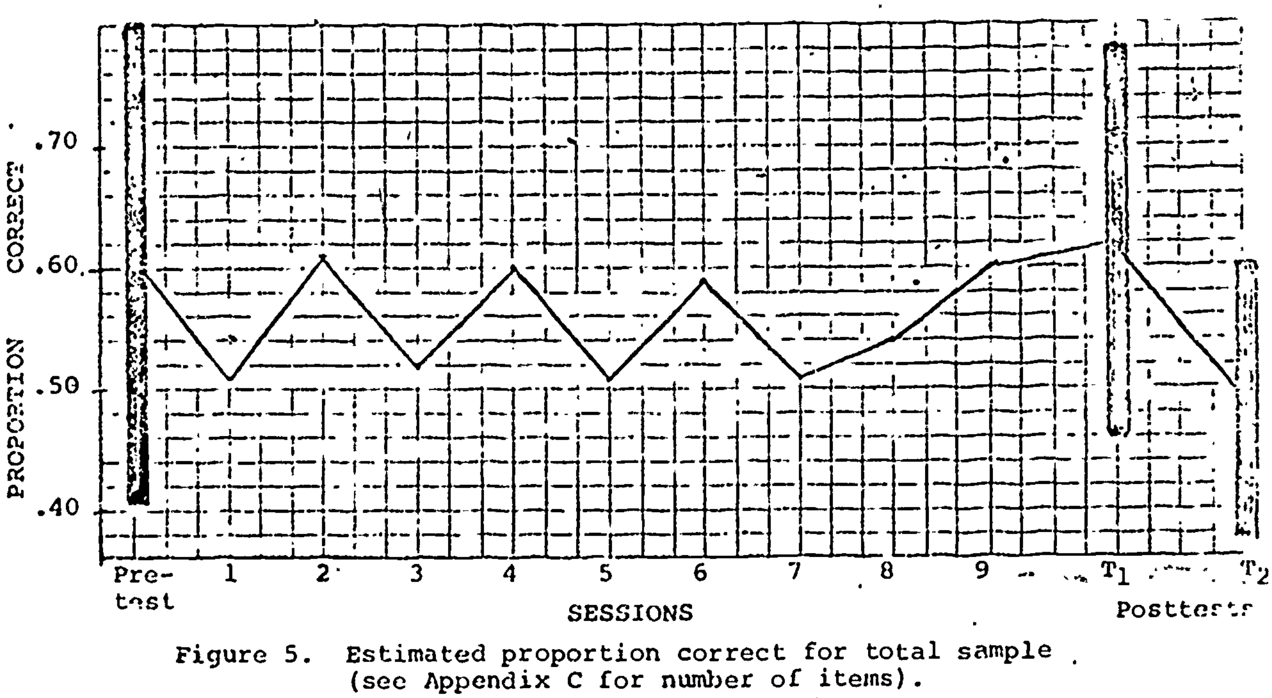

Figure 5. Estimated proportion correct for total sample (see Appendix C for number of items).

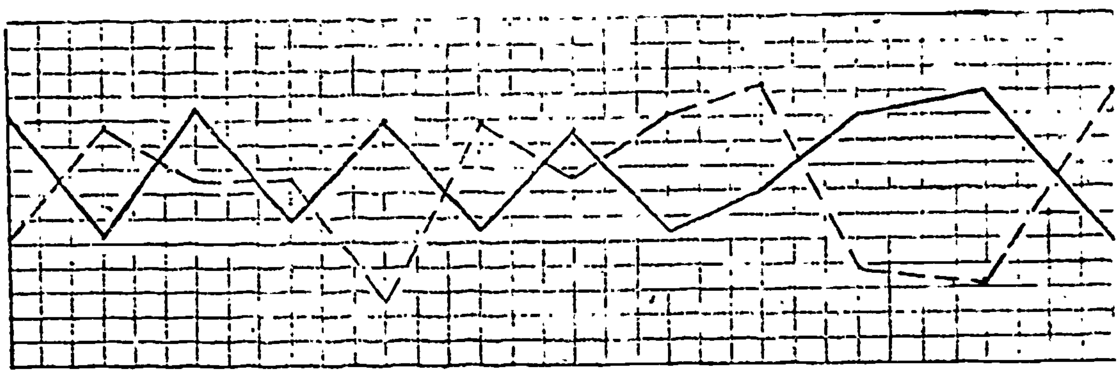

Figure 6. Relationship between realism (broken line) and difficulty (solid line). See Figures 4 and 5.

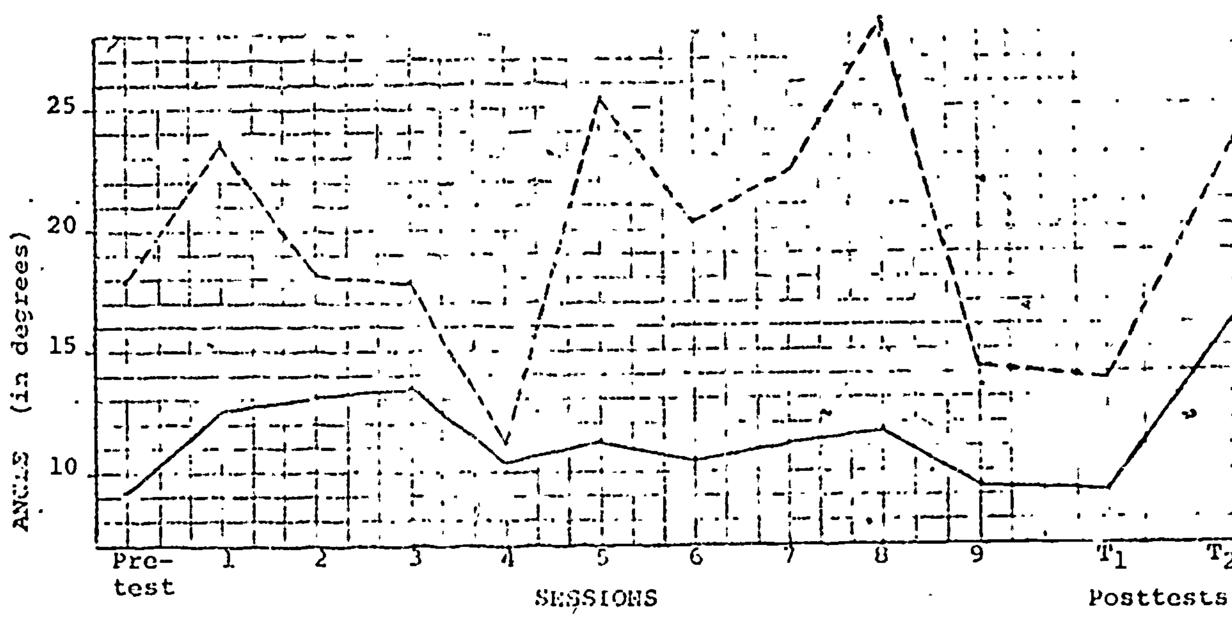

Figure 7. Angle between ideal and observed realism for high class (solid line) and low class (broken line).

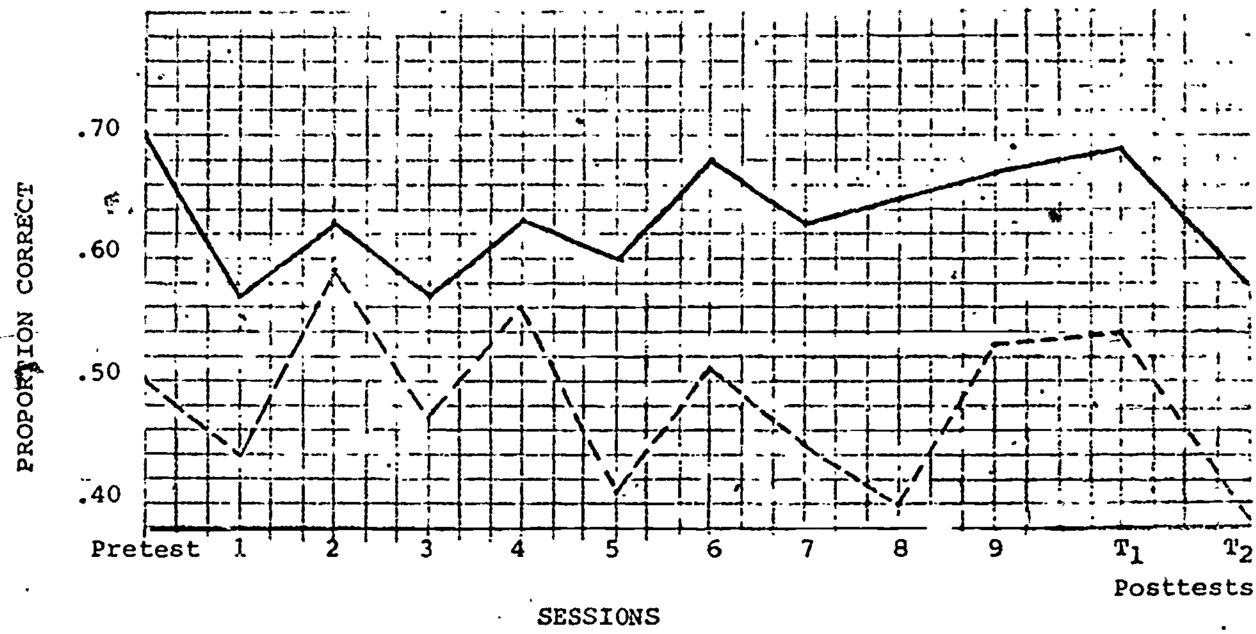

Figure 8. Estimated proportion correct for high class (solid line) and low class (broken line).

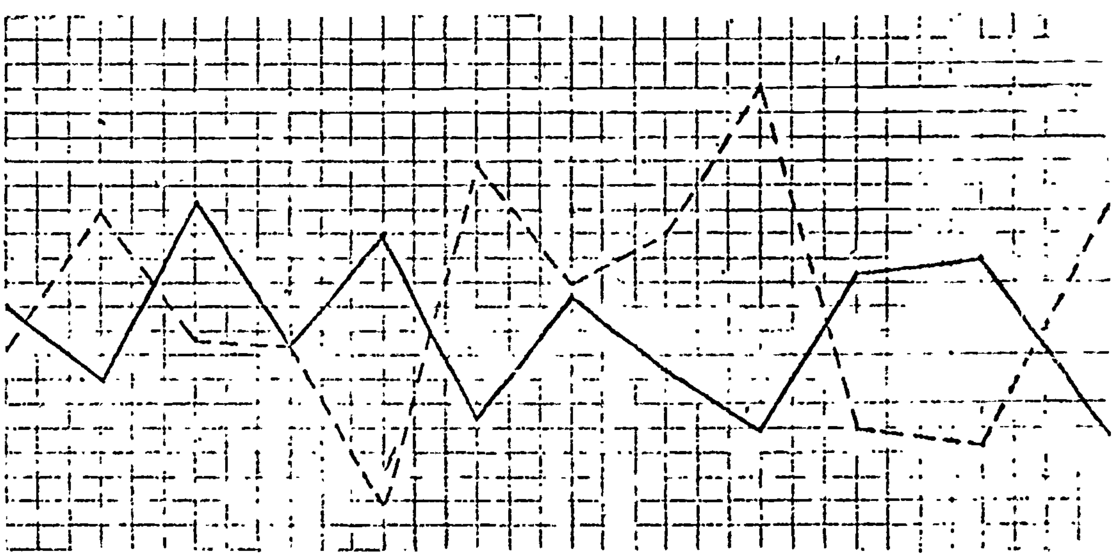

Figure 9. Relationship between realism (broken line) and difficulty (solid line) for low class. See Figures 7 and 8.

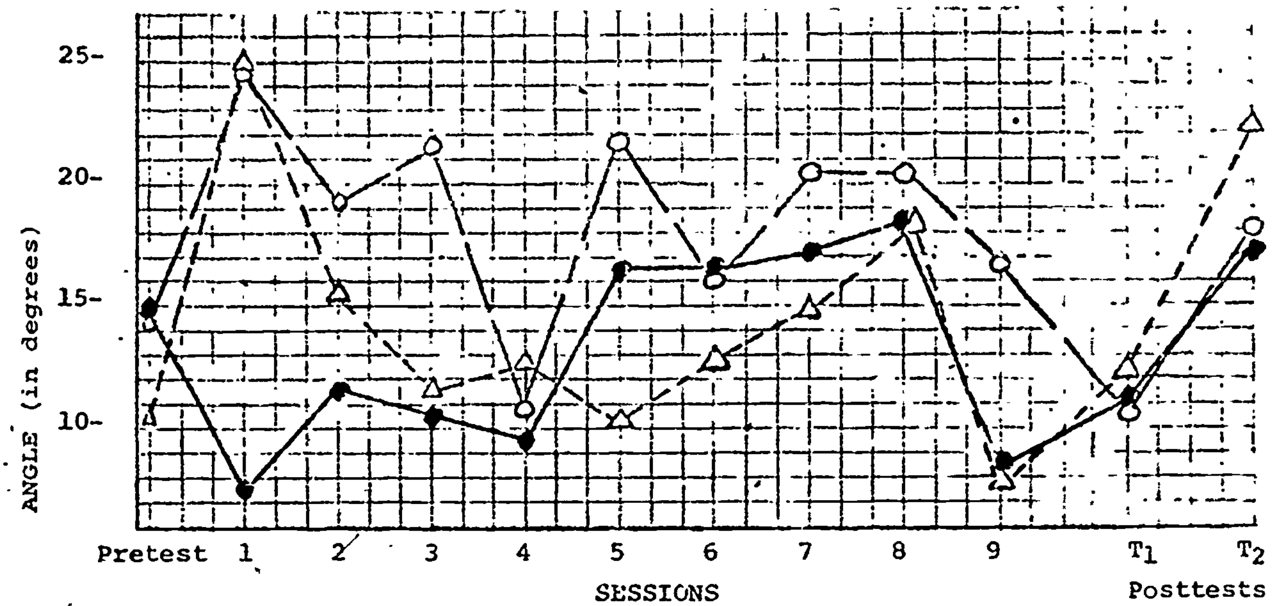

Figure 10. Angle between ideal and observed realism for treatment groups (1=closed circle, 2= open circle, 3=triangle).

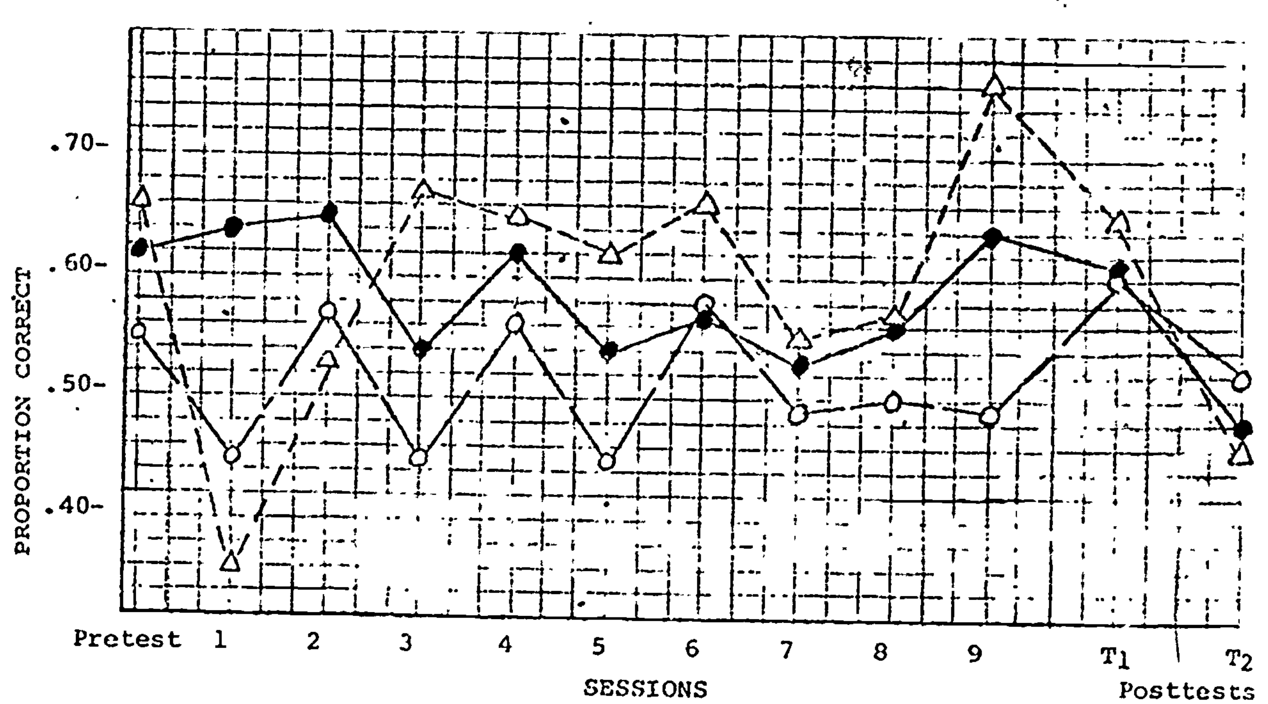

Figure 11. Estimated proportion correct for treatment groups (1=closed circle, 2=open circle, 3=triangle).

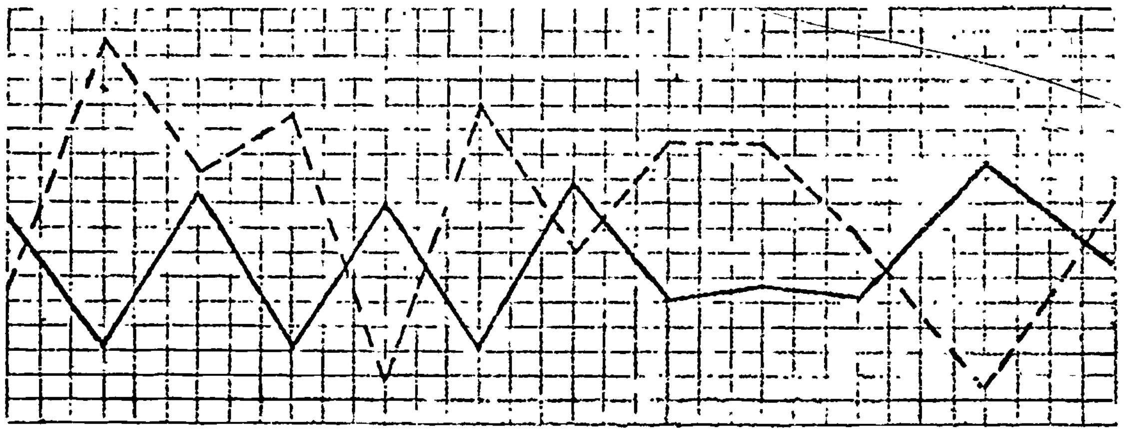

Figure 12. Relationship between realism (broken line) and difficulty (solid line) for treatment group receiving individual study with the realism graph (see Figures 10 and 11).

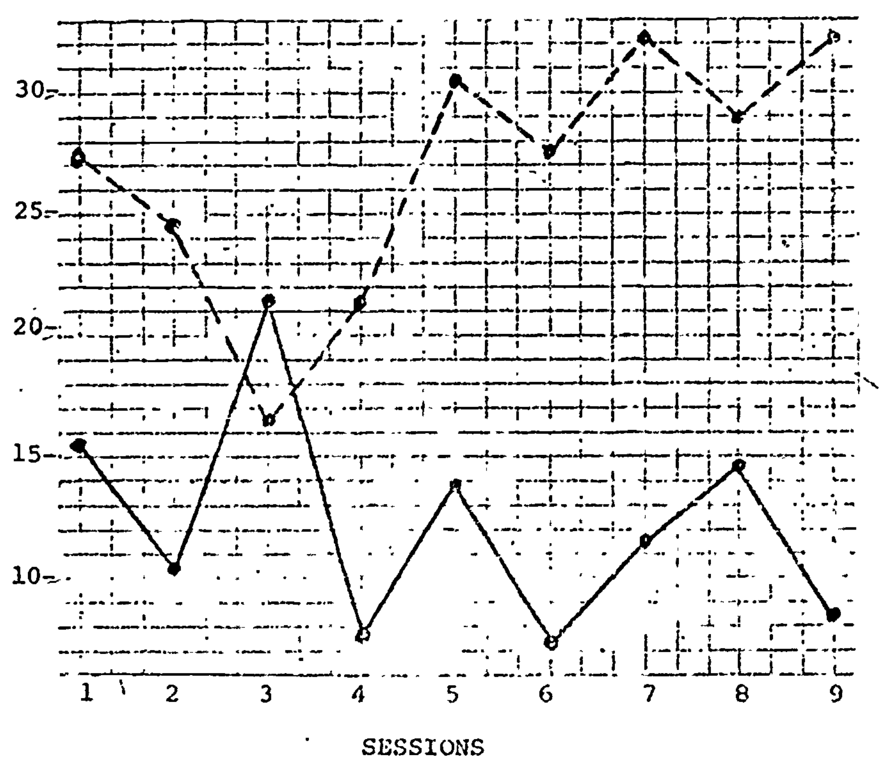

Figure 13. Angle between ideal and observed realism for easy items (solid line) and hard items (broken line) during training (see Appendix D for Ns and mean number of items).

Realism gains. Despite wide variations in performance from session to session, Figures 4, 7, and 10 all suggest a developing tendency toward improved realism. The overall standard deviation on the first posttest was smaller than on the pretest, with the mean angle of Session 9 at the same level as the posttest (Figure 4). The two classes were closer in realism performance by the time of the first posttest (Figure 7), and treatment groups tended to converge over the training sessions (Figure 10).

Although the larger variability and generally larger realism angles in the early sessions may suggest that Ss were less familiar with the SCoRule than later in the experiment, the proportion of Ss having realism angles within five degrees of perfect realism was actually largest in the first session (Figure 14). Moreover, the proportion was relatively constant except for sessions 6 and 8. When the proportion is examined by classes, the curve for the low class tends to rise, but the curve for the high class tends to fall (Figure 15). While the two classes finished with a similar proportion attaining defined realism, it was because the low class gained.

The three treatment groups differed widely in the proportion attaining defined realism (Figure 16). The two individual study groups tended to move in opposite directions, the one without the graph having a higher proportion in the early sessions and the one with the graph performing more realistically in later sessions and on the first posttest. The curves of these two groups bear similarities with those of the high and low classes, especially with respect to the crossover that occurred midway through the experiment. Figure 17 breaks apart the curves for the individual study treatment groups in order to compare high and low class proportions in these treatments. Except for the sharp jump on the first posttest (no S was "realistic" on the second posttest), the low group with the graph exhibited little variability but upward spurts began to appear after the first few training sessions. The spurts resembled the more pronounced variation in the high group with the graph, which also appeared late in training.

The four groups of Figure 17 appear in Figure 18 to show that the pattern toward realism is not so apparent when the mean angle of deviation from perfect realism is considered. Some relative gain in realism for graph groups over no-graph groups is suggested by the downward slope of the high graph group and the catch-up pattern of the low graph group, but angle of deviation was not as indicative of improvement as proportion attaining defined realism.

Overconfidence

The direction of deviation from realism was almost

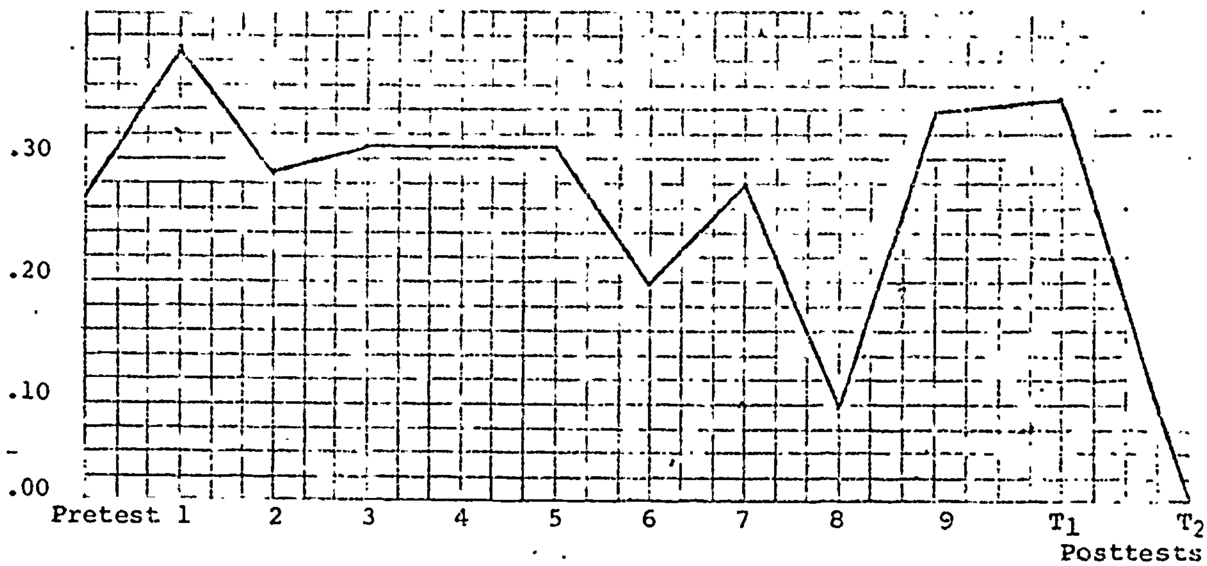

Figure 14. Proportion of total sample whose observed realism was within five degrees of perfect realism.

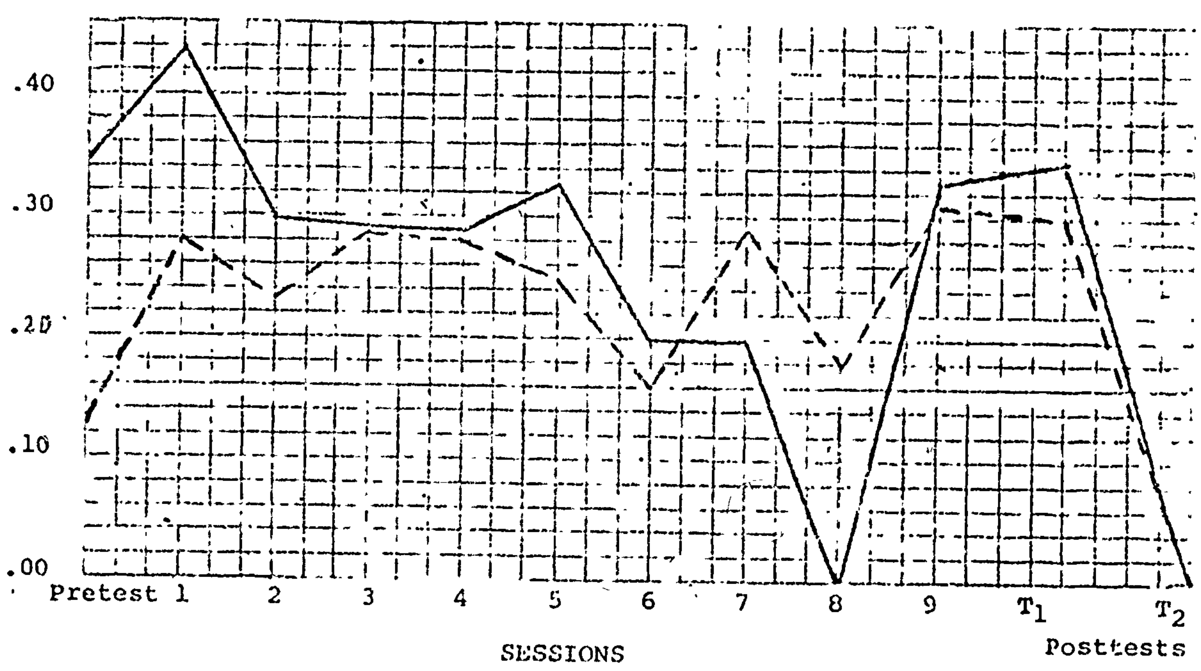

Figure 15. Proportion of high class (solid line) and low class (broken line) whose observed realism was within five degrees of perfect realism.

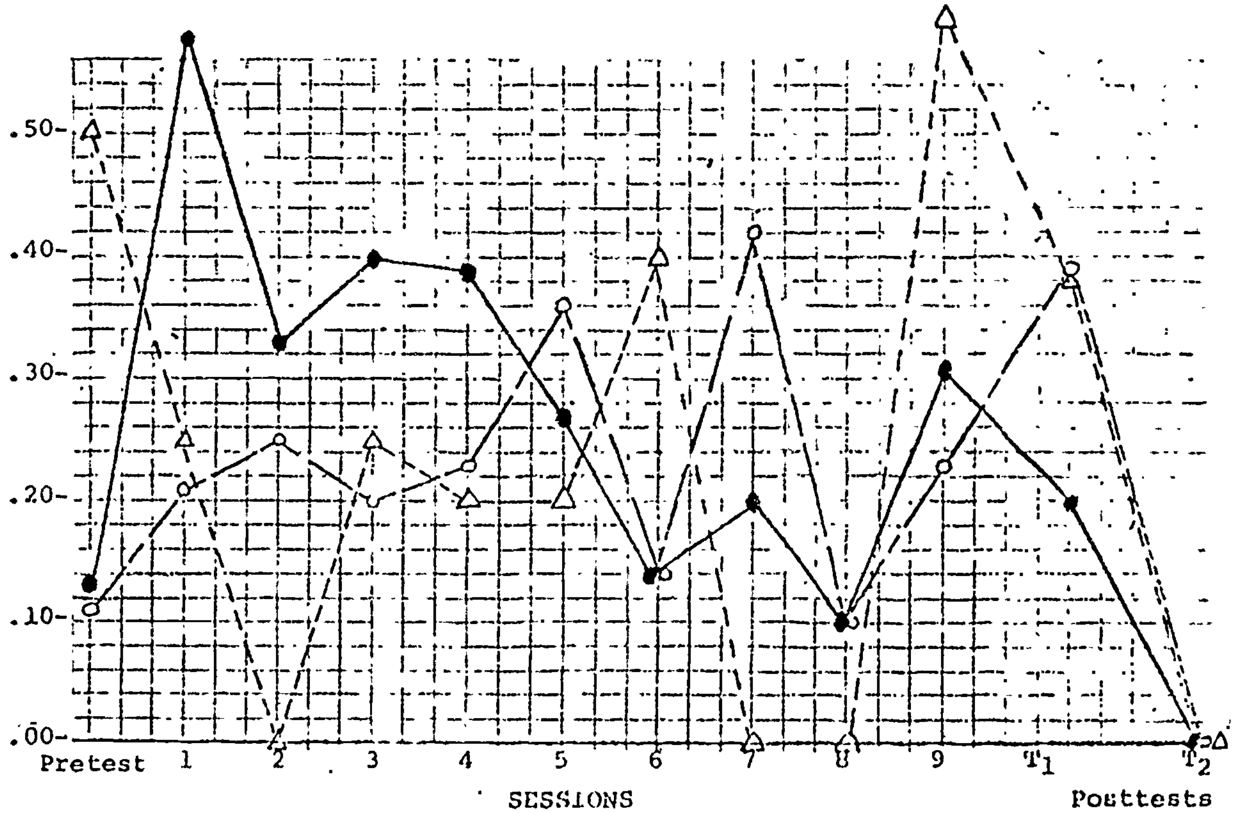

Figure 16. Proportion of treatment group 1 (solid circles), 2 (open circles), and 3 (triangles) whose observed realism was within five degrees of perfect realism.

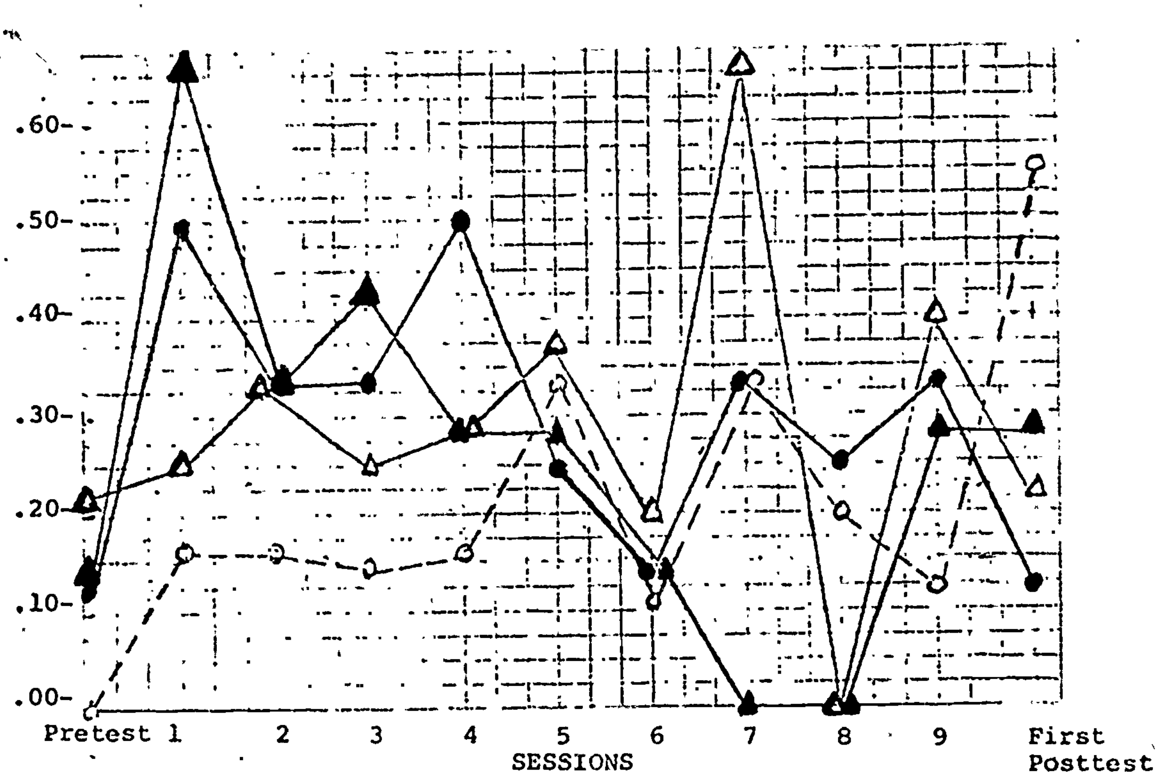

Figure 17. Proportion attaining defined realism in the individual study treatments: low-class-graph (open circle), low-class-no graph (closed circle), high-class graph (open triangle), and high-class-no graph (closed triangle).

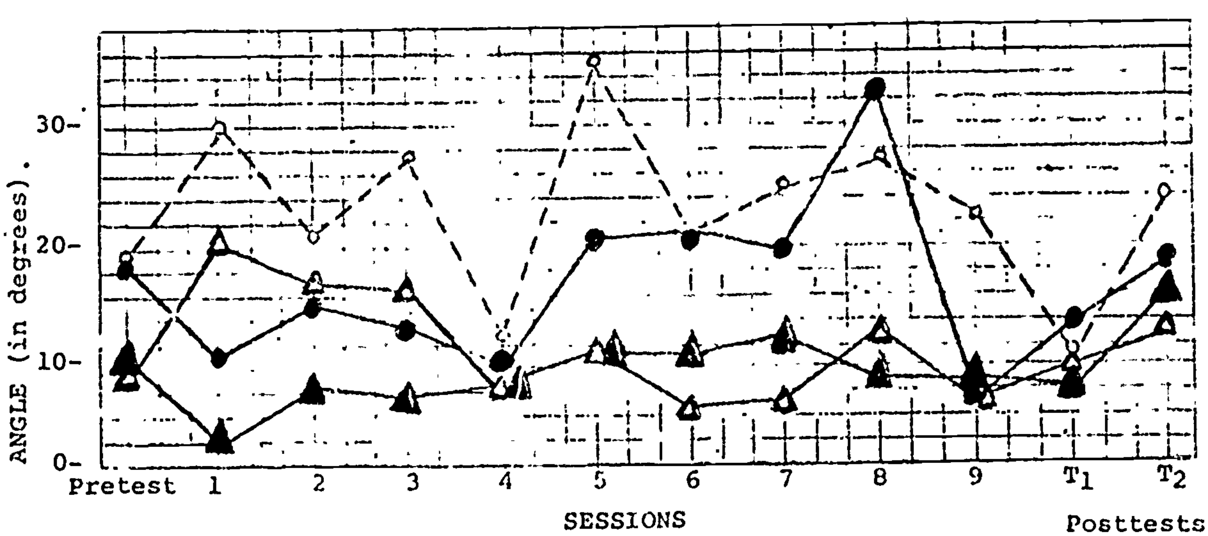

Figure 18. Angle between ideal and observed realism for the four groups of Figure 17.

uniformly toward overconfidence throughout the experiment (Table 8). Of those with unrealistic behavior, only four tended toward underconfidence on the pretest, only three on the first posttest and only two on the second posttest. The proportions during training sessions were approximately the same as on the tests. The sample showed a strong tendency to overvalue information, i.e. to assign too much weight to what was mistakenly believed to be the correct alternative. Deviation from realism was especially strong on the second posttest, nearly all of it toward overconfidence.

The bulk of those who were underconfident during the experiment were in the low class (Table 9). Of the three treatments the individual study group with the graph tended to do the most undervaluing of information (Table 10). The least undervaluing occurred in the team study group, in which all lack of realism was in the direction of overconfidence. Team testing (posttest 2) was also marked by little undervaluing despite the strong deviation from realism.

In general, few Ss were consistently able to refrain from temptations to mass their probabilities on the wrong response. The overall pattern indicates that Ss commonly placed more value on their knowledge and their reasoning than their performance justified. Since lack of realism was nearly always in the direction of overconfidence, both realism and overconfidence appear to be sensitive to problem difficulty. Overconfidence increased along with difficulty in nearly all cases.

One exception is the tendency of the low class to have a higher proportion of underconfident Ss than the high class, even though the high class found the problems less difficult. A similar exception is the contrast between treatment groups: the group which had the most difficulty with the items (individual study with graph) had the highest proportion of underconfident Ss. These exceptions suggest that underconfidence and overconfidence were both sensitive to problem difficulty but the higher the level of achievement the more likely S deviated from realism in the direction of overconfidence.

Another factor contributing to overconfidence appeared to be team discussion. Without exception, any team that deviated from realism during training sessions exhibited overconfidence. No team member in Group 3 showed underconfidence on any of the tests, and only two Ss from the entire sample undervalued information on the teamed posttest.

IV. Discussion

Although evidence from this experiment fails to confirm the general effectiveness of realism training in modifying assessment tendencies, the study produced several findings about assessment tendencies of the seventh graders who

Table 8.
Number of Ss in Three Categories of Observed Realism
(Total Sample)

	Underconfident	Realistic*	Overconfident
Pretest	4	12	33
Session			
1	2	11	17
2	4	7	15
3	2	10	22
4	2	9	20
5	0	10	24
6	2	6	25
7	0	7	20
8	1	2	22
9	3	10	18
Posttest 1	3	16	30
Posttest 2	2	0	40

*realism angle within 5 degrees of perfect realism

Table 9.
Number of Ss in Three Categories of Observed Realism
By Class

	Underconfident		"Realistic"*		Overconfident	
	High	Low	High	Low	High	Low
Pretest	1	3	9	3	16	17
Session 1	1	1	7	4	8	9
2	1	3	4	3	9	7
3	1	1	5	5	11	11
4	0	2	5	4	12	8
5	0	0	6	4	12	12
6	0	2	3	3	12	13
7	0	0	2	5	8	12
8	0	1	0	2	14	8
9	0	3	5	5	10	8
Posttest 1	0	3	9	7	17	13
Posttest 2	0	2	0	0	24	16

*observed realism within 5 degrees of perfect realism

Table 10.
Number of Ss in Three Categories of Observed Realism By Treatment Groups

	Underconfident			"Realistic"*			Overconfident		
	1**	2	3***	1	2	3	1	2	3
Pretest	2	2	0	2	2	8	11	14	8
Session									
1	0	2	0	7	3	1	5	9	3
2	1	3	0	4	3	0	7	6	2
3	0	2	0	6	3	1	9	12	3
4	1	1	0	5	3	1	7	11	4
5	0	0	0	4	5	1	11	9	4
6	0	2	0	2	2	2	12	10	3
7	0	0	0	2	5	0	8	7	5
8	0	1	0	1	1	0	9	8	5
9	2	1	0	4	3	3	7	9	2
Posttest 1	2	1	0	3	7	6	10	11	10
Posttest 2	1	1	0	0	0	0	13	12	15

*observed realism within 5 degrees of perfect realism
**1=individual study, no graph; 2=individual study, graph; 3=team study, graph
***The unit of analysis for Group 3 during sessions is the team.

participated in the experiment. In addition, the data indicated that realism training was more effective for low achieving students, and a pattern emerged among low achievers favoring individual study with a realism graph as feedback over no graph or team study. In view of the circumstances of the experiment, these findings suggest that a more extensive training program is needed before drawing the conclusion that realism cannot be substantially modified.

Earlier work of an unsystematic nature has indicated that assessment tendencies of junior high students appear to be stable over time and may not respond to training. In this study, the proportion of students whose observed realism was within five degrees of perfect realism increased from pretest to the first posttest by only 8%. However, every sub-grouping of the sample showed a decrease in the standard deviation of the major dependent variable (angle of deviation from perfect realism). Certain groups, such as the low-achieving class and the individual-study treatments, responded more favorably to training than other groups. In some cases, notably individual study with the realism graph as feedback, a distinct trend toward realism occurred over the training sessions.

Student Characteristics

The effect of training was not as evident from this study as several characteristics of students' assessment tendencies. First, the tendencies of some students were much more stable than those of others. The low class in particular fluctuated widely in marked contrast to the steady behavior of the high class. Second and most prominent in the findings, students were characteristically overconfident in their use of information. The slope of observed realism nearly always fell below perfect realism when the angle of deviation exceeded five degrees. The tendency to place more value on an alternative than one's information about that alternative is actually worth appears to be relatively constant in the data. Despite increased opportunities to use the SCoRule and gain more familiarity with the process of assessing subjective probabilities, students overvalued information toward the end of the experiment in about the same way they did at the beginning whenever they were not being realistic.

Third, students became less realistic as problems became more difficult. The wide fluctuations in observed realism that occurred in some of the data are regularly related to changes in difficulty level. The relationship, however, does not appear to be linear because the size of the fluctuations was much greater when problems were especially difficult (e.g. the low class) than when difficulty was reduced (e.g. the high class).

Fourth, the data suggests that difficulty level may be related to overconfidence. In groups for which problems were relatively less difficult, nearly all students who failed to achieve defined realism exhibited overconfidence, while most of those few who were underconfident were in groups that had relatively more difficulty with the problems.

Finally, overconfidence may also be related to team discussion. The absence of underconfidence in the team study group and the nearly unanimous expression of overconfidence on the team posttest may indicate that the dynamics of group interaction promote a tendency to ignore the limitations of one's information. The group, for example, may function to divert an individual from recognizing what he knows by subtly undermining his respect for his own judgment or by introducing extraneous personality factors into the decision-making process. The practice of assigning students to congenial team groupings may have contributed to a tendency to adopt the position of peers despite misgivings.

It should be noted that the overconfidence so characteristic of this sample is not to be confused with an exaggerated sense of self-worth or an emotional state. Overconfidence, as Figure 2 indicates, is a cognitive construct dealing with use of information. It may be that affective characteristics play a major role in how information is assessed, but the primary meaning of overconfidence should be interpreted in the context of decision theory (de Finetti, 1937; Raiffa, 1968; Ramsey, 1931; Savage, 1954). It is when a student attaches value to his knowledge and reasoning about a problem in excess of the value his knowledge and reasoning are really worth that he becomes overconfident. It is altogether possible that a student has a weak self-concept or appears unsure of himself while his observed realism indicates overconfidence, since the distortion reflected in his deviation from realism is oriented toward information rather than toward interpersonal relations.

At the same time it must be noted that a relationship between overconfidence and personality characteristics is at least suggested by the data. Risk-taking tendencies and peer interaction are both important ingredients in decision theory, especially in relation to group problem-solving. If difficulty increases to a point at which the individual feels personally threatened, information distortion is likely to be in part a function of self-esteem. Whether he will express overconfidence is problematic, however, because individuals differ. Some react more cautiously as difficulty mounts, e.g. the underconfident students in the low class. Others become overconfident, refusing to acknowledge that they know less than they did on the easier problems. Self-concept may interact with assessment tendencies even more when interpersonal relations are part of the decision process. Members with weak self-concepts may acquiesce while their partners interpret acquiescence as a signal to weight suggestions more strongly than they should. Information use is distinct from affective characteristics but probably closely related.

Realism Training

Besides the general reduction in the standard deviation of information distortion (angle of deviation from perfect realism), the data supports a tentative conclusion that realism training has beneficial effects in some instances despite the lack of dramatic gains overall. The low class made steady gains in the proportion of students who were able to achieve observed realism within five degrees of perfect realism. In later training sessions, this class even surpassed the high class, and on the first posttest showed a gain while the high class proportion remained at the pretest level (Figure 15).

The source of this gain within the low class came exclusively from the second treatment group, in which individual study was combined with the realism graph. The sharp increase is even more striking in view of the lack of gain in the other five cells of Table 7. Although neither the proportion attaining defined realism nor the angle of deviation from perfect realism indicated dramatic progress during the training sessions, the graph group receiving individual study in the low class did appear to make spurts in the direction of improvement (Figures 17 and 18).

Realism gain may require a combination of feedback about realism and high problem difficulty. The apparently strong relationship between difficulty and realism may mean that students are more likely to discover their long-established tendencies to distort the value of information when they confront a high level of uncertainty than when they have to deal with uncertainty that is only moderate. The discovery, moreover, may require explicit feedback about their tendencies in order to establish an intuitive awareness of the subjective responses that contribute to realistic behavior.

If explicit feedback is important to attaining realism, it may seem strange that the angle of deviation from perfect realism fails to distingui sharply between the two individual study groups in the low class. It should be remembered, however, that explicit feedback was given only with respect to observed realism within five degrees of perfect realism. The proportion attaining defined realism is then the best index of the effectiveness of explicit feedback, and the data do support the use of a realism graph for students having relative difficulty with the problems.

While the lack of overall gains may be related to inappropriate difficulty levels or the absence of explicit feedback in some cells of the design, a more likely explanation is the brevity of the training. The modification of long-established tendencies probably requires a period of months. It is possible that something of the effect of a longer period may be achieved in just three weeks if the training is concentrated, but the evidence of this study indicates that nine 20-25 minute sessions are not enough. In fact, it is surprising that an average of 7-8 items per session provided enough training to be associated with any change in observed realism, e.g. the reduced standard deviations and the modest gain in the number attaining defined realism.

Time constraints also weakened the effect of the realism graph, because some students were not able to stay long enough at the end of a session to view their graphs. The rush to board the bus tended to interfere with the process of receiving explicit feedback both in terms of viewing the graph and listening to experimenter interpretations of the graph. With full and frequent exposure to realism graphs over several months, students may be able to develop assessment strategies that substantially alter established tendencies.

Contrary to expectations, team problem-solving did not appear to support realism training. Especially when working with relatively difficult problems (second posttest), students found realism hard to achieve while interacting with their peers. One of the questions raised by this experiment is why team study is so closely associated with overconfidence.

Even without team study, however, seventh graders clearly tend to exaggerate the value of their information. This key finding agrees with Savage's observation, quoted earlier, that our culture, especially our schooling, encourages us to confuse belief with certainty. Though we have no assurance that realism can be taught, it would seem that programs for realism training merit further study.

REFERENCES

Brennan, R. Manual for DEC-TEST: A FORTRAN IV computer program for decision-theoretic test scoring and the analysis of item data. Addendum to Final Report for Grant OEG-2-2-2B118, Office of Education, U.S. Department of Health, Education, and Welfare, Washington, D.C. 1973.

deFinetti, B. Foresight, its logical laws, its subjective sources. In H. Kyburg, Jr., and H. Smokler (eds.), _Studies in subjective probability_, New York; Wiley, 1964 (Translated by Kyburg from the French of 1937).

Raiffa, H. _Decision analysis: Introductory lectures on choices under uncertainty_. Reading, Mass.: Addison-Wesley, 1968.

Ramsey, F. _The foundations of mathematics and other logical essays_. Condon: Rutledge and Kegan Paul, 931.

Savage, L. _The foundations of statistics_. New York: Wiley, 1954.

Savage, L. Elicitation of personal probabilities and expectations. _Journal of the American Statistical Association_, 66, 1971.

Shuford, E. SCoRule response aid instructions. Lexington, Mass.: SM Corporation, 1970.

Shuford, E. Potential application of admissible probability measurement to elementary and secondary education. Research proposal (mimeo). Santa Monica, Calif.: Rand Corporation, 1973.

Shuford E. Albert A. & Massengill, A. Admissible probability measurement procedures. _Psychometrika_, 31, 1966.

APPENDIX A.

SCoRule Procedures Test*

NAME______________________________ TEACHER________________

	A	B	C	D
a.	76	76	60	60
b.	70	70	70	70
c.	95	66	0	0
d.	100	0	0	0
e.	95	43	43	43
f.	85	85	0	0

Which of the above SCoRule settings would be best if...

__d__ 1. you are sure that A is the correct choice. (Really sure!)

__c__ 2. you are almost sure that A is correct but B might be correct. (You know that C and D are wrong.)

__a__ 3. you are not sure which choice is correct but you have reason to think that A or B are more likely to be correct than C or D.

__b__ 4. you have no knowledge to help you make a choice.

__c__ 5. you are almost sure that A is correct but if it isn't you have no idea which of the others is.

__b__ 6. you will just have to guess.

(Write a <u>little</u> letter in each of the blanks.)

*Scoring: 2 points per question, 1 point for "f" on question 3.

APPENDIX B.

Computational Procedures for Calculating Realism during Training Sessions*

1. Student's responses = S(X) = S(A), S(B), S(C), S(D)**

   a. Each X is a one to three digit number ($0 \leq X \leq 100$) which is input separately.
   b. Each X is a SCoRule value (log score).
   c. Order of input proceeds from first value (A) thru last value (D).

2. Transformation from log score, S(X), to probability, p(X).

   a. $0 \leq p(X) \leq 1.00$
   b. $p(X) = \text{anti}\log_{10}\left(\frac{S(X) - 100}{50}\right)$

3. Validity check on responses.

   a. $\left|1.00 - \sum_X p(X)\right| \leq \Delta$

   b. $\Delta = .04$***

4. Preparing student's response, p(X), for curve fitting.

   a. $d(X) = \begin{cases} 1 - p(X) & \text{if X is correct answer} \\ - p(X) & \text{if X is incorrect answer} \end{cases}$

   b. $PD \equiv \sum_X p(X)\, d(X)$

   c. $P^2 \equiv \sum_X p(X)^2$

**d. $a_i \equiv$ number of possible answer to ith item = 4

5. Accumulating response data over items

   a. $K \equiv$ number of items answered in the session

   b. $\sum_1^k PD$ and $\sum_1^k P^2$ are sums accumulated item by item, i.e. $\sum_1^i PD = \sum_1^{i-1} PD + PD$; is computed at the end of item i.

**c. $\sum_1^k a_i = K * 4$

6. Curve fitting at end of session

**a. $\bar{p} = \frac{K}{\sum a_i} = .25$

b. slope $b = \frac{\sum PD}{\sum P^2 - K\bar{p}}$

7. Category assignment for feedback.

a. A slope of .82 is five degrees below b=1.00.

b. A slope of 1.20 is five degrees above b=1.00.

c. "Realistic" if $.82 \leq b \leq 1.20$.

d. "Overconfident" if $b < .82$.

e. "Underconfident" if $b > 1.20$.

*Computational procedures for DEC-TEST, which was used to analyze the data, are presented in Brennan (1973).

**Only four alternative items were used in this experiment.

***This tolerance level was arbitrary; it should not be confused with the range of tolerance for defined realism (7 above).

APPENDIX C.

Summary Data for Tests
And Training Sessions

NOTE: Unless otherwise indicated, the standard deviation is shown below the mean, and the number of Ss and mean number of items are hyphenated on the line below the standard deviation.

Example:	8.32	($\overline{X}$)
	6.51	(sd)
	13 - 7.2	(N-$\overline{K}$)

C-1 Cell Means for Angles of Deviation of Observed Realism from Perfect Realism.

	Low Class Treatment Group 1	Low Class Treatment Group 2	Low Class Treatment Group 3	High Class Treatment Group 1	High Class Treatment Group 2	High Class Treatment Group 3	TOTAL SAMPLE
Pretest*	18.47 12.93 8-18.8	19.05 9.05 9-19.0	15.76 17.14 6-22.3	10.84 6.26 7-21.6	9.67 8.48 9-18.9	7.54 7.10 10-23.1	13.31 11.34 49-20.6
Sessions							
1	10.98 13.45 6-5.5	30.11 27.43 6-5.7	42.37 16.18 2-8.5	4.00 5.65 6-4.8	20.34 21.27 8-5.6	7.55 7.47 2-6.0	17.77 21.48 30-5.7
2	15.24 18.58 6-8.0	21.23 20.84 6-7.3	15.76 .00 1-4.0	8.46 6.85 6-6.3	20.72 17.15 6-5.2	15.77 .00 1-5.0	15.54 17.59 26-6.5
3	13.27 10.98 8-8.8	27.30 20.30 7-6.4	3.80 1.37 2-9.5	7.84 8.60 7-5.0	16.63 13.71 8-5.8	19.74 .01 2-6.0	15.65 14.97 34-6.7
4	10.14 11.70 6-8.7	12.77 10.51 6-6.2	9.13 2.07 2-6.0	9.35 7.87 7-4.6	9.30 5.41 7-4.9	15.31 11.12 3-7.0	10.72 9.20 31-6.1
5	21.10 18.73 8-10.1	35.00 32.18 6-7.5	13.60 2.16 2-8.5	11.62 11.38 7-8.1	11.89 9.66 8-7.4	7.93 5.77 3-11.0	17.83 19.94 34-8.6
6	21.70 22.75 7-8.7	21.21 18.34 9-7.9	10.36 10.34 2-5.5	11.40 4.88 7-9.6	6.75 3.12 5-8.6	14.44 8.31 3-9.0	15.77 15.96 33-8.5
7	19.94 21.09 6-6.2	24.68 27.39 9-6.4	18.98 10.46 2-7.0	12.76 6.39 4-9.0	7.59 10.57 3-9.0	12.25 3.95 3-10.0	18.16 20.32 27-7.5
8	32.50 23.99 4-6.3	26.97 26.51 5-6.2	25.16 13.96 2-6.5	9.17 3.42 6-10.5	13.65 3.47 5-9.4	14.01 1.38 3-8.3	19.22 18.18 25-8.2
9	7.52 5.14 6-7.7	22.59 22.36 8-7.5	.65 .64 2-5.0	9.00 7.41 7-9.1	7.71 6.39 5-9.6	12.85 9.55 3-8.7	11.85 14.43 31-8.2
Posttests							
1	13.57 8.11 8-32.1	11.41 11.52 9-27.6	17.83 10.48 6-34.3	8.22 4.51 7-36.4	10.13 5.73 9-29.3	8.89 6.66 10-32.7	11.34 8.68 49-31.8
2	18.69 8.09 7-20.0	26.11 5.05 5-15.8	26.98 13.17 6-19.8	16.06 2.81 7-17.4	13.04 3.92 8-18.9	19.04 3.86 9-20.3	19.32 8.33 42-18.9

* About five minutes of the 45 minute class period was spent completing the SCoRules Procedures Test.

C-2. Marginal Means for Angles of Deviation of Observed Realism from Perfect Realism

	Class		Treatment Group			TOTAL SAMPLE
	Low	High	1	2	3	
Pretest*	17.99	9.17	14.91	14.36	10.62	13.31
	12.98	7.53	11.04	9.94	12.55	11.34
	23-19.8	26-21.2	15-20.1	18-18.9	16-22.8	49-20.6
Sessions						
1	23.67	12.61	7.49	24.52	24.96	17.77
	23.96	17.50	10.89	24.59	21.49	21.48
	14-6.0	16-5.4	12-5.2	14-5.6	4-7.3	30-5.7
2	18.04	13.03	11.85	19.19	15.77	15.54
	19.20	15.42	14.41	20.88	.01	17.59
	13-7.4	13-5.7	12-7.2	12-6.3	2-4.5	26-6.5
3	17.93	13.37	10.73	21.61	11.77	15.65
	17.22	11.89	10.30	17.91	8.03	14.97
	17-7.9	17-5.5	15-70	15-6.7	4-7.8	34-6.7
4	11.12	10.38	9.71	10.90	12.84	10.72
	10.43	8.04	9.83	8.35	9.23	9.20
	14-7.2	17-5.1	13-6.5	13-5.5	5-6.6.	31-6.1
5	25.38	11.12	16.68	21.80	10.20	17.83
	25.01	9.97	16.43	25.06	5.44	19.94
	16-8.9	18-8.3	15-9.2	14-7.4	5-10.0	34-8.6
6	20.20	10.46	16.55	16.05	12.81	15.77
	19.84	6.03	17.24	16.36	9.39	15.96
	18-7.9	15-9.1	14-9.1	14-8.1	5-7.6	33-8.5
7	22.34	11.06	17.07	20.41	14.94	18.16
	23.94	7.73	17.20	25.40	8.00	20.32
	11-6.4	10-9.3	10-7.3	12-7.1	5-8.8	27-7.5
8	28.65	11.81	18.50	20.31	18.47	19.22
	23.94	3.87	19.18	20.04	10.43	18.18
	11-6.3	14-9.6	10-8.8	10-7.8	5-7.6	25-8.2
9	14.20	9.34	8.32	16.86	7.97	11.85
	18.30	7.80	6.51	19.38	9.52	14.43
	16-7.3	15-9.2	13-8.5	13-8.3	5-7.2	31-8.2
Posttests						
1	13.84	9.14	11.08	10.77	12.24	11.34
	10.48	5.88	7.19	9.13	9.36	8.68
	23-30.9	26-32.5	15-34.1	18-28.4	16-33.3	49-31.8
2	23.52	16.17	17.37	18.07	22.22	19.32
	10.26	4.40	6.20	7.73	9.67	8.33
	18-18.8	24-19.0	14-18.7	13-17.7	15-20.1	42-18.9

*About five minutes of the 45-minute class period was spent completing the SCoRule Procedures Test.

C-3 Cell Means and Standard Deviations for Estimated Proportion Correct*

	Low Class Treatment Group 1	2	3	High Class Treatment Group 1	2	3	TOTAL
Pretest	.53 .18	.46 .19	.53 .22	.73 .07	.64 .18	.74 .15	.61 .20
Sessions							
1	.54 .16	.39 .22	.26 .09	.73 .14	.49 .28	.46 .11	.51 .24
2	.57 .22	.61 .31	.56 .00	.73 1	.54 .25	.50 .00	.61 .24
3	.44 .16	.41 .18	.79 .09	.66 .14	.49 .22	.55 .05	.52 .20
4	.55 .17	.53 .08	.70 .05	.69 .16	.58 .26	.61 .21	.60 .19
5	.41 .17	.37 .19	.55 .02	.68 .19	.50 .14	.67 .08	.51 .20
6	.46 .19	.52 .20	.68 .18	.69 .10	.70 .12	.65 .13	.59 .19
7	.47 .15	.45 .27	.41 .16	.63 .11	.60 .23	.64 .05	.51 .22
8	.33 .28	.43 .31	.45 .13	.71 .11	.57 .14	.64 .04	.54 .23
9	.54 .21	.44 .21	.85 .15	.72 .18	.58 .17	.70 .14	.60 .23
Posttests							
1	.53 .12	.56 .18	.52 .14	.70 .07	.64 .16	.72 .13	.62 .16
2	.42 .09	.41 .05	.35 .06	.55 .08	.59 .08	.54 .07	.49 .11

* See Appendix C-1 for N and $\overline{K}$ in each cell.

C-4 Marginal Means and Standard Deviations for Estimated Proportion Correct*

	Class		Treatment Group			TOTAL
	Low	High	1	2	3	SAMPLE
Pretest	.50 .20	.70 .15	.62 .18	.55 .21	.66 .20	.61 .20
Sessions						
1	.44 .21	.57 .25	.64 .18	.45 .26	.36 .14	.51 .24
2	.59 .26	.63 .21	.65 .19	.57 .28	.53 .03	.61 .24
3	.47 .20	.57 .19	.54 .19	.45 .21	.67 .14	.52 .20
4	.56 .14	.63 .22	.62 .18	.56 .20	.65 .17	.60 .19
5.	.41 .18	.60 .18	.54 .23	.45 .18	.62 .09	.51 .20
6	.51 .20	.68 .11	.57 .19	.58 .19	.66 .15	.59 .19
7	.45 .23	.63 .14	.53 .16	.49 .27	.55 .16	.51 .22
8	.40 .26	.65 .13	.56 .25	.50 .25	.57 .13	.54 .23
9	.53 .24	.67 .18	.76 .16	.64 .21	.60 .73	.60 .23
Posttests						
1	.54 .15	.69 .13	.61 .13	.60 .17	.65 .16	.62 .16
2	.39 .08	.56 .08	.48 .11	.52 .11	.46 .12	.49 .11

* See Appendix C-2 for N and $\overline{K}$ information.

C-5. Cell Means and Standard Deviations for Mean Log Scores (Average SCoRule Value for Correct Alternatives)*

	Low Class Treatment Group 1	2	3	High Class Treatment Group 1	2	3	TOTAL SAMPLE
<u>Pretest</u>	70.06	71.23	73.92	78.02	75,68	80.76	75.10
	9.58	7.33	8.74	3.64	10.44	11.63	9.93
<u>Sessions</u>							
1	75.49	58.13	63.17	83.54	64.10	76.73	69.85
	8.10	17.48	1.17	9.95	22.58	6.12	17.86
	.0**	.5	.0	.7	.5	.5	.4
2	73.98	77.58	67.50	80.62	70.05	68.00	74.96
	13.11	13.13	.00	11.70	11.88	.00	12.74
	1.2	1.0	.0	.3	1.5	.0	.9
3	70.72	71.68	85.12	79.74	68.54	61.43	72.56
	8.59	5.63	2.39	11.11	17.91	1.43	12.54
	1.1	.7	1.0	.0	.6	.0	.6
4	80.14	71.85	77.71	77.08	77.13	70.79	76.10
	6.82	6.54	3.71	12.50	9.27	17.83	10.60
	.3	.8	.0	.0	.7	.0	.4
5	65.96	65.90	71.10	77.23	69.25	80.10	70.59
	12.81	12.83	3.32	10.83	9.51	8.97	12.13
	1.2	1.0	2.5	.4	.4	.0	.8
6	68.27	71.49	76.10	74.74	79.73	70.71	72.95
	7.21	8.14	17.90	8.02	5.65	13.52	9.88
	.7	.4	3.0	.1	.2	.0	.5
7	68.02	69.74	74.77	70.06	77.09	70.58	70.69
	14.61	19.86	2.23	11.29	15.72	5.47	15.38
	.0	.2	2.0	.0	.0	.0	.2
8	53.67	69.92	59.89	78.10	71.56	70.33	68.86
	17.16	21.11	13.44	5.96	5.40	2.87	15.22
	.0	.8	2.5	1.0	.0	.0	.6
9	78.93	63.02	87.19	79.20	76.53	72.42	74.40
	9.67	24.88	12.81	12.05	7.91	15.93	17.61
	.3	.3	2.5	.6	.0	.0	.4
<u>Posttests</u>							
1	69.87	75.17	66.19	78.61	76.64	79.10	74.77
	7.47	9.76	11.91	6.83	7.26	11.37	10.31
2	66.45	58.49	57.57	67.16	70.59	62.38	64.27
	4.61	9.09	12.39	3.50	4.74	5.52	8.31

*See Appendix C-1 for N and $\overline{K}$ in each cell.
**Mean number of computer validity checks, i.e. mean number of times in which an item's probability values (anti log of SCoRule values) did not sum to 1.00 within a .04 tolerance. <u>S</u> then re-entered values.

C-6. Marginal Means and Standard Deviations for Mean Log Scores*

	Class			Treatment Group		TOTAL SAMPLE
	Low	High				
Pretest	71.52 8.67	78.26 9.90	73.78 8.42	73.45 9.29	78.19 11.14	75.10 9.93
Sessions						
1	66.29 15.02 .2**	72.97 19.49 .6	79.51 9.93 .3	61.54 20.76 .5	69.95 8.09 .3	69.85 17.86 .4
2	75.14 12.91 1.0	74.77 12.57 .9	77.30 12.86 .8	73.81 13.07 1.3	67.75 .25 .0	74.96 12.74 .9
3	72.81 8.30 .9	72.32 15.67 .3	74.93 10.83 .6	70.01 13.73 .7	73.28 12.01 .5	72.56 12.54 .6
4	76.24 7.44 .5	75.99 12.71 .3	78.49 10.39 .2	74.69 8.54 .8	73.56 14.41 .0	76.10 10.66 .4
5	66.58 12.17 1.3	74.16 10.93 .3	71.22 8.28 .4	67.82 11.18 .6	76.50 8.50 1.0	70.59 12.13 .8
6	70.75 9.73 .8	75.60 9.38 .1	71.51 8.28 .4	74.43 8.34 .4	72.87 15.65 1.2	72.95 9.88 .5
7	69.72 16.99 .4	72.33 12.00 .0	68.84 13.42 .0	71.57 19.17 .2	72.26 4.91 .8	70.69 15.38 .2
8	62.18 19.92 .8	74.10 6.29 .4	68.32 16.80 .6	70.74 15.43 .4	66.16 10.17 1.0	68.86 15.22 .6
9	72.01 21.27 .6	76.96 12.08 .3	79.07 11.02 .5	68.23 21.18 .2	78.33 16.44 1.0	74.40 17.61 .4
Posttests						
1	70.99 10.35	78.12 9.04	73.95 8.40	75.91 8.64	74.26 13.16	74.77 10.31
2	61.28 9.98	66.51 5.88	66.80 4.11	65.93 8.96	60.46 9.23	64.27 8.31

*See Appendix C-2 for N and $\overline{K}$ information

**Mean number of computer validity checks (see Appendix C-5).

APPENDIX D.

Means and Standard Deviations for Angles of Deviation
of Observed Realism from Perfect Realism
On Easy and Hard Items
During Training Sessions

Session	Easy			Hard		
	N	$\overline{K}$*	$\overline{X}$*	N	$\overline{K}$	$\overline{X}$
1	26	2.96 1.22	15.58 18.62	25	3.12 1.17	27.26 26.87
2	26	3.35 1.41	10.29 14.49	26	3.46 1.17	24.53 23.80
3	28	3.50 1.37	21.55 25.36	32	3.53 1.11	16.30 19.97
4	25	3.32 1.22	7.69 11.95	29	3.59 1.18	21.19 21.48
5	35	4.26 1.50	13.92 19.87	33	4.36 1.45	30.67 26.69
6	36	4.06 1.33	7.42 9.70	32	4.34 1.43	27.67 22.21
7	25	3.72 1.40	11.62 19.48	25	3.92 1.44	32.25 24.76
8	25	4.24 1.51	14.81 13.85	24	3.63 1.41	29.08 23.71
9	34	3.97 1.61	8.38 15.65	32	4.13 1.43	32.13 27.69

*Standard deviations are shown below means.