DOCUMENT RESUME

ED 032 221 SE 007 479

By-Klopfer, Leopold E.

An Evaluative Study of the Effectiveness and Effects of Astronomy Materials Prepared by the University of Illinois Elementary-School Science Project.

Chicago Univ., Ill. Graduate School of Education.

Pub Date [64]

Note-59p.

EDRS Price MF-$0.50 HC-$3.05

Descriptors-*Astronomy, Earth Science, *Elementary School Science, *Evaluation, Instructional Materials, *Science Course Improvement Project

Identifiers-National Science Foundation, Test on Understanding Science

Evaluated was the effectiveness of the materials of one book, "Charting the Universe," of the six books that comprise the University of Illinois Elementary Science Project. Five hypotheses were tested, including one related to students' general understanding of science, and another related to students' views of astronomy, arithmetic, scientists, and learning experiences in science. Instruments used were Test on Understanding Science (TOUS), two locally constructed subject-matter achievement tests, and an experimental designed semantic differential instrument to measure children's perceptions of science. The student population (43 boys and 49 girls) consisted of the entire fifth grade in the University of Chicago Laboratory School during the school year 1963-64. All students were taught for ten weeks by the same person, a science teacher at the Laboratory School. Major findings of the study were: (1) students were moderately successful in mastering some of the topics taught; (2) students' general knowledge of astronomy increased during the ten weeks of instruction; (3) the effect of studying these materials on general understanding of science were slight; (4) studying the materials did affect the students' view of astronomy, but did not affect their view of learning experiences in science. (BR)

AN EVALUATIVE STUDY OF THE EFFECTIVENESS AND EFFECTS OF ASTRONOMY MATERIALS PREPARED BY THE UNIVERSITY OF ILLINOIS ELEMENTARY-SCHOOL SCIENCE PROJECT

LEOPOLD E. KLOPFER
Assistant Professor of Education
in the Natural Sciences
Graduate School of Education
The University of Chicago

Participating Teacher
for the Study:

Barbara Wehr
The University of Chicago
Laboratory Schools

Research Assistants
for the Study:

Fred Geis, Jr.
E. Lawrence Liss
Mary E. McCullough

INTRODUCTION

So long as there are schools and so long as there is some dissatisfaction with what children study in schools, new curriculum materials will be developed. Perhaps because there are now so many schools in the United States and so much dissatisfaction with what children study in them, the development of new curriculum materials is proceeding today at a previously unprecedented pace. Amidst the flurry of often richly endowed curriculum development activity, the conviction is steadily gaining ground among educators that the curriculum materials produced are not necessarily "good" simply because they are "new." Before foisting the new products of curriculum development projects on unsuspecting children, responsible educators are asking pertinent questions about the outcomes that can be expected from using the new materials and their suitability for different groups of students. Fortunately, the curriculum developers, by and large, have accepted the responsibility for seeking answers to such question as a part of their developmental work.

When he considers the outcomes to be anticipated from students' use of his materials, two kinds of questions confront the curriculum developer. The first concerns the effectiveness of the materials in getting students to learn the particular subject matter that they are designed to teach. Second, and at least equally important, is the effect of the materials on the students' general perceptions of the subject or discipline being studied and of its modes of inquiry. The study presented here considers aspects of both of these kinds of questions of concern to the curriculum developer. It seeks to illustrate some of the ways by which a relatively modest evaluative study can be of considerable benefit in furthering the work of a curriculum development project.

One of the shortcomings of many evaluative studies of curriculum materials has been that the approach has been too gross. Typically, total test scores are used to measure student achievement or pretest-posttest changes in mean scores are used to measure student gain. While these measures are valuable and should be a necessary part of an evaluative study, they fail to provide the curriculum developer with sufficient specific information about the strengths and weaknesses of his materials. Particularly in the early phases of a curriculum development project, it is important to have as specific information as possible about what knowledge and which ideas are mastered successfully by the students, about where the students failed to attain mastery, and about the changes, if any, in students' perception of the subject that accompany instruction with the new materials. Data that will yield such specific information can be obtained quite readily in a carefully designed evaluative study. In fact, much of the data of this kind is frequently collected in the course of a study, but the data are seldom fully exploited in the analysis. This study illustrates some procedures of analysis and interpretation that may be utilized to yield information of direct value for the continuing development of curriculum materials.

DESCRIPTION OF THE STUDY

Curriculum materials developed by the University of Illinois Elementary School Science Project (ESSP) were the subject of this study. The focus of the ESSP is the development of curriculum materials in the area of astronomy - "materials that are sound astromically, that reflect the structure of the subject as it is viewed by astronomers of stature, and that can be handled by teachers and children in actual classrooms."[1] The ESSP materials for students consist of a series of six booklets, richly illustrated with line drawings and containing reading text and many appropriately interspersed pupil activities. A comprehensive Teacher's Guide accompanies each of the student booklets. Charting the Universe (1963 edition), Book 1 of the series, was used in this study.

Purposes of the Study

Two main purposes were conceived for this study, *viz.*, (A) to assess the effectiveness of the ESSP materials; and (B) to assess the effect on students of studying the ESSP materials. (In the preceding sentence, and throughout this report, the term "ESSP materials" should be understood as referring only to ESSP Book 1, Charting the Universe.) Under each of these purposes, several related questions were given consideration.

---

[1] J. Myron Atkin, "Some Evaluation Problems in a Course Content Improvement Project," *Journal of Research in Science Teaching*, **1**, 129-132 (1963).

Assessment of the effectiveness of the ESSP materials included an attempt to ascertain how well the students learned the particular topics the materials were designed to teach. The Teacher's Guide (page i) states that "Book I presents a sequential development of ideas to show how astronomers are able to chart the universe," and this sequence of ideas represents the topics to be mastered through study of the materials. Moreover, the ESSP curriculum developers and the investigator believed that this approach to the study of astronomy would also result in a concomittant increase in students' general knowledge of astonomy, even though such specific information about astronomy was not explicitly taught in Book 1. Hence, the assessment of subject matter achievement included both of these aspects. The hypothesis tested was:

HYPOTHESIS 1. Study of ESSP materials will increase students' knowledge of astronomy and of how astronomical information is obtained.

Implicit in this hypothesis, in view of the emphasis in certain sections of Book 1, is that "how astronomical information is obtained" includes the mastery of several skills in making measurement. A further question related to this first hypothesis regarding subject-matter achievement concerned the connection between any such achievement and students' general scholastic ability. For investigating this question, the hypothesis formulated was:

HYPOTHESIS 2. Subject matter achievement is positively correlated with a student's general scholastic ability.

The second main purpose of this study was to assess the effect on students of studying the ESSP materials. Questions under this head are concerned with investigating changes in students' perceptions of certain aspects of science and science study. Though the effecting of such changes in perceptions may not be an explicitly stated objective, the influence of curriculum materials on students' perceptions is an inevitable accompaniment of instruction. We believed it desirable and important to investigate this effect of the ESSP materials. Specifically investigated were possible changes in students' general understanding of science, the relationship of any such changes to subject-matter achievement, and possible changes in students' perception of astronomy, arithmetic, scientists, and the study of science. The pertinent hypotheses were:

HYPOTHESIS 3. Study of ESSP materials will increase students' general understanding of science (as measured by the Test On Understanding Science).

HYPOTHESIS 4. A student's gain in general understanding of science is positively correlated with subject-matter achievement.

HYPOTHESIS 5. Study of ESSP materials will affect students' views of astronomy, arithmetic, scientists, and learning experiences in science.

Instruments

To obtain data for assessing the effectiveness of the ESSP materials, two subject-matter achievement tests were constructed. Some of the multiple-choice items for these tests were obtained from tests used previously by the ESSP, but most of the test items were devised especially for this study. The subject-matter pretest (called "Charting the Universe Test, Form 207") consisted of 28 multiple-choice items. Of these, 15 items dealt with material specifically taught in ESSP Book 1, and 13 items tested for selected topics of general knowledge in astronomy. (The Book 1 items on the subject-matter pretest did not touch on all the material taught in Book 1; in order not to make the test too long and to avoid a possibly frustrating experience for the children, no questions were included on Book 1 material about which none or very few of the pupils were expected to know prior to the study. The participating teacher and the investigator together decided what topics should not be included.) From the pretest administration, the test reliability of the total test was found to be .597 (Kuder-Richardson Formula 20). The reliability for the subtest of Book 1 items was .353, and the reliability of the subtest of General Knowledge items was .475.

The subject-matter posttest (called "Charting the Universe Test, Form 208") contained 42 items and included the 28 items from the pretest, nine additional multiple-choice items concerned with Book 1 material, and five items calling for the student to demonstrate his skills in making measurements of lines and angles as taught in Book 1. For the total test, the reliability computed from the posttest administration data was .829. For the subtest consisting of the 28 pretest items the reliability was .676; for the subtest of the Book 1 pretest items, the reliability was .569; for the subtest of General Knowledge items, the reliability was .603; and for the subtest of all Book 1 items, the reliability was .759.

To obtain data bearing on the study's second main purpose, to assess the effect on students of studying the ESSP materials, two additional instruments were administered both as pretest and posttest. One of these was the Test On Understanding Science (TOUS), Form Ex, which is one of a group of instruments, developed by the investigator and several collaborators, to measure students' understanding of salient aspects of the aims and processes of scientific enquiry, the characteristics of scientists, and the dynamics of the scientific enterprise. TOUS, Form Ex, contains 36 multiple-choice items, many of which call for the making of quite careful discriminations to select the best answer from the four alternative responses presented. From the pretest administration of TOUS in this study, the test reliability computed was .578; from the posttest administration, the reliability was .643. (The values for the test reliability of TOUS are somewhat lower than those typically found from other studies.)

Lastly, to complete the testing battery for this study, an experimental semantic differential instrument was designed. The semantic differential developed by Osgood and his associates, though hitherto little used by researchers in science education, provides a promising technique for assessing students' perceptions of concepts relevant to the teaching of science. In a typical semantic differential instrument, the student is asked to indicate his associations of a given concept with a series of bipolar word-pairs (e.g., good-bad, powerful-weak, exciting-dull). Working rapidly, he checks the one of five or more available positions between each pair of bipolar adjectives which represents how he associates these words with the concept. The result of the checking process is a series of ratings of the given concept along a dozen or more adjectival bipolar scales. The same set of scales is usually used for rating several concepts appearing on successive pages of the semantic differential instrument.[2] This was the practice adopted in the semantic differential instrument, called "Word Association Study" (WAS), designed for the present study.

The WAS instrument consisted of two cover pages containing an explanation of how to make responses on it and eight pages of 15 five-position adjective scales to be used in rating the following eight concepts: ASTRONOMY, ARITHMETIC, MOST SCIENTISTS, EXPLORING NEW IDEAS, DOING SCIENCE EXPERIMENTS, READING ABOUT SCIENCE, MAKING MEASUREMENTS, MY SCIENCE TEACHER. The 15 bipolar adjectival pairs, in the order of their appearance on each page, were: quick-slow, weak-powerful, dirty-clean, hard-soft, important-unimportant, dull-exciting, mannish-womanish, good-bad, unenjoyable-enjoyable, moving-still, useless-useful, changing-permanent, foolish-wise, interesting-boring, easy-difficult. Eight concepts times 15 scales gives a total of 120 ratings to be made by a student on the WAS instrument. The use of the WAS instrument in this study represents the first application, as far as we know, of the semantic differential technique in an evaluative study of elementary-school science curriculum materials.

---

2 For further discussion of the theory and development of the semantic differential, see Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum, The Measurement of Meaning, University of Illinois Press, Urbana, Ill., 1957.

h

DOING SCIENCE EXPERIMENTS

quick	A	B	C	D	E	slow
weak	A	B	C	D	E	powerful
dirty	A	B	C	D	E	clean
hard	A	B	C	D	E	soft
important	A	B	C	D	E	unimportant
dull	A	B	C	D	E	exciting
mannish	A	B	C	D	E	womanish
good	A	B	C	D	E	bad
unenjoyable	A	B	C	D	E	enjoyable
moving	A	B	C	D	E	still
useless	A	B	C	D	E	useful
changing	A	B	C	D	E	permanent
foolish	A	B	C	D	E	wise
interesting	A	B	C	D	E	boring
easy	A	B	C	D	E	difficult

Sample page from
WAS instrument

Population and Procedures

The students included in the study comprised the entire fifth grade in the University of Chicago Laboratory School during the school year 1963-64. These 92 students, 43 boys and 49 girls, were in four instructional groups of 23 students each. The range of I.Q. scores (Henmon-Nelson Test: Elementary Form) for the entire group was 88 to 179, with a median score of 124.

All of the groups were taught by Miss Barbara Wehr, science teacher in the University of Chicago Laboratory Schools. Miss Wehr has had more than ten years of teaching experience with a strong emphasis in elementary-school science. For ten weeks of instruction, the pupils studied the ESSP materials in Book 1, Charting the Universe. Each group met for three 50-minute periods per week. A copy of the pupil book was provided for each child, and the suggested equipment and supplies for all the pupil exercises were made available. The teacher carefully followed the "suggestions for teaching" presented in the Teacher's Guide and also chose to include most of the "supplementary activities" and "supplementary exercises." In the course of the instruction, she prepared 11 sheets of additional exercise material for the use of the students.

The three pretests (Charting the Universe Test, Form 207; TOUS; and WAS) were administered to the pupils in the four instructional groups on 7, 10, and 11 February 1964. None of the pretests was administered by the participating teacher.

Study of the ESSP materials began in each instructional group on the class meeting following the third pretest and continued in every succeeding class meeting during ten weeks of instruction. The instructional period was interrupted by one week of vacation and one week of no science classes during a school camping trip. Following the instructional period, the three posttests (Charting the Universe Test, Form 208; TOUS; and WAS) were administered on 8, 9, and 10 May.

In constituting the four instructional groups at the beginning of the school year, no selection criteria had been applied and pupils were randomly assigned to a group. During the instructional period of this study, all four groups used the same materials and were taught with the same procedures by the same teacher. Hence, the four instructional groups were considered to be a single population, and the individual student was taken as the unit of analysis. Data collected in the study were punched into IBM cards, and data processing was accomplished through the facilities of the Education Statistics Laboratory and the IBM 7094 Computation Center at the University of Chicago.

FINDINGS

A. EFFECTIVENESS OF THE ESSP MATERIALS

Subject-Matter Achievement

In order to test the first hypothesis, that study of ESSP materials will increase students' knowledge of astronomy and of how astronomical information is obtained, a *t* test was used comparing the scores on the pretests with the scores on the same measures as posttests. The *t* was computed for a Total Astronomy Test, which was the combined scores of the General Knowledge and Book 1 items, and for the General Knowledge Test and for the Book 1 Test separately. Because the pre and post scores were obtained from the same pupils, they were presumed to be related, and the *t* for correlated groups was computed. The results are shown in Table 1.

The posttest results, in all three cases indicate that a significant difference exists. More than chance factors were involved in the increase, and the greater achievement can probably be ascribed to the use of the ESSP materials.

To test the second hypothesis, that subject matter achievement is positively correlated with a student's general scholastic ability, a partial correlation coefficient was derived. Performance on the posttest was assumed to be related to performance on the pretest. Our purpose, however, was to detect the correlation between posttest scores and I.Q., nullifying the effect on the pretest. A first order partial correlation provides this information.

TABLE 1

Comparison of Pretest and Posttest Scores on Subject Matter Tests
(N = 90)

		Mean	S.D.	t	Significance Level
Book 1	pre	5.59	2.088	7.63	$p < .001$
	post	7.36	2.555		
General Knowledge	pre	4.82	2.12	3.84	$p < .001$
	post	5.62	2.41		
Total	pre	10.41	3.52	7.58	$p < .001$
	post	12.98	4.47		

Partial correlation coefficients were obtained for the General Knowledge items, the Book 1 items, and for a total score which combined the first two. (See Table 2). Transforming the partial r to a corresponding Z value, we found a confidence level for the value of the partial correlation using the normal distribution of the statistic $Z(-4)^{\frac{1}{2}}$, according to the procedure described by Hays.[3]

As can be seen from the table, all three correlations are significant. Both on items particular to the ESSP materials and on items measuring general knowledge about astronomy, achievement was related positively, though only slightly, to general scholastic ability.

As with any correlation coefficient, care must be exercized in the interpretation. Using the relationship that r^2 equals the proportion of the total variance of one factor accounted for by the other, we find that only about 7% of the variance of Book 1 scores is accountable to I.Q. Only about 8% of the variance of General Knowledge scores is attributable to I.Q.

Intelligence tests tend to be tests of verbal ability. This may be what we are attempting to correlate with our achievement test scores. We find little relation because the ESSP materials seem not to demand such verbal facility, and success with them is more likely to be related to other factors we have not measured.

---

3 William L. Hays, Statistics for Psychologists. Holt, Rinehart and Winston, New York, 1963. Page 576.

TABLE 2

Partial Correlation Coefficients: I.Q. with Subject-Matter Posttest Holding the Pretest Constant

	I.Q.	Significance Level
Book 1	.260	$p < .01$
General Knowledge	.294	$p < .01$
Total	.308	$p < .01$

Analysis of Test Items

While the analysis and comparison of whole-test means yields information as to the effectiveness of the methods and/or materials being used, closer inspection of specific items may reveal areas of knowledge and understanding in which the materials are particularly successful or unsuccessful.

A technique for this kind of analysis is McNemar's chi square test of change.[4] A fourfold contingency table is constructed for each item illustrating the numbers who had the item right or wrong on the protest and posttest. Chi square

		Pretest	
		Wrong	Right
Posttest	Right	A	B
	Wrong	C	D

is equal to $(A-D)^2/(A+D)$ and has one degree of freedom. Using this statistic, we obtain a measure of the significance of the change in the responses to the item.

Of the 28 test items which appear on both the pretest and the posttest, 11 had significant changes. These 11 were comprised of 6 of the 13 general knowledge items and 5 of the 15 Book 1 items.

---

4 Quinn McNemar, Psychological Statistics. John Wiley, New York, 1962. Pages 224-227.

Book 1 Items: Table 3

Three of the Book 1 items with significant changes from pre to posttest are all concerned with the topic of angles and measurement of angles and triangles(#7, #14, and #22). Not only did more pupils choose the right answer on the posttest, but also, except in the case of choice A in item #22, fewer pupils chose each incorrect answer. Since these three items are the total number of items on the test measuring achievement of knowledge on this topic, it appears that the materials are quite effective in helping children learn about the measurement of angles and triangles. Among the pupils, 72% got item #7 correct on the posttest; 86% got item #14 correct, and 63% got item #22 correct. The first two can certainly be considered to indicate mastery of the materials, and the latter approaches mastery if we consider a 70% class achievement to be our criterion.

Two other Book 1 questions are included in Table 3. On both of these items also, the change from pretest to posttest results in a significant chi square. But on the posttest only 15% of the pupils got #21 correct; apparently the students did not learn to do the kind of estimating called for in this test item. Only 22% of the pupils got #28 correct. The most popular incorrect response to this item both on the pretest (55%) and the posttest (42%) was alternative C, which names the circle which is the largest as drawn in the diagram. The idea of apparent angular diameter was probably not adequately mastered by the students. Of note also, among 92 pupils, 74 got #21 wrong

TABLE 3

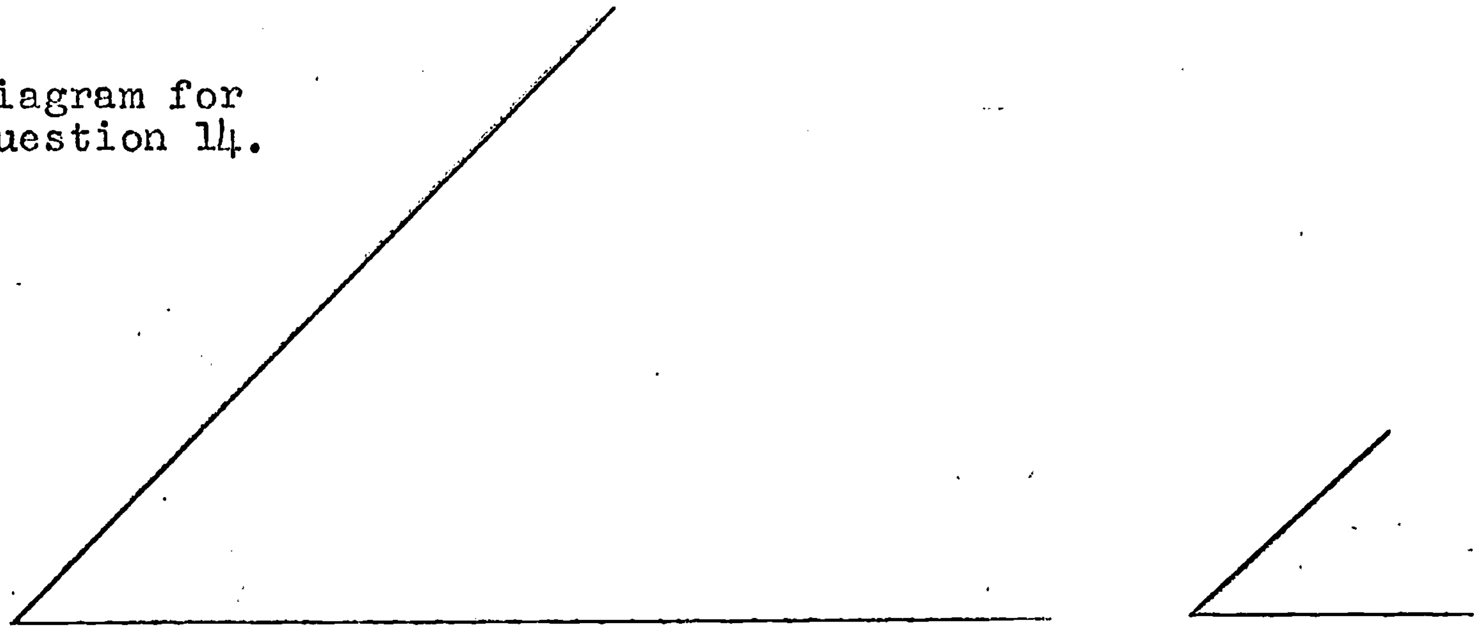
Book 1 Items with Significant Chi Square Test of Change:
Proportions of Responses by Students

7. One angle of a triangle is 40 degrees and another one is 70 degrees. What must the third angle be?

- A. It could be anything; triangles come in all sizes.
- *B. 70 degrees
- C. 40 degrees
- D. You need to measure to find out.
- E. $57\frac{1}{2}$ degrees

	A	B	C	D	E	X^2	
pre	.27	.30	.11	.20	.08	27.769	$p < .001$
post	.08	.72	.04	.12	.04		

Diagram for Question 14.



14. Of L and M above, which is the larger angle?

- A. Angle L. It covers more of the page.
- B. Angle M.
- C. It depends on what you mean by "angle."
- *D. It looks like they are both the same, but you need to measure to be sure.
- E. It depends on what size circle they are in.

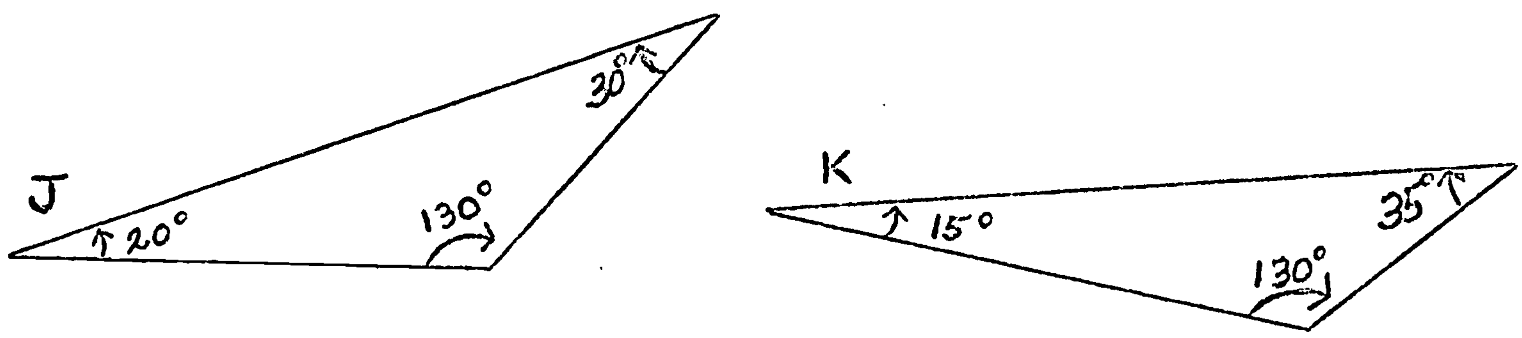
	A	B	C	D	E	X^2	
pre	.15	.07	.08	.57	.14	17.780	$p < .001$
post	.05	.05	.02	.86	.01		

TABLE 3 (cont'd)

21. Tom made several measurements with his ruler. One of the lines he measured ended between the mark indicating 8 inches and the mark indicating 8 1/10 inches. It was slightly nearer the 8 inch mark. Which of the following numbers should Tom record, if he wants to record the most accurate measurement?

A. 8 inches.
*B. 8.0 inches.
C. 8.00 inches.
D. 8.1 inches.

	A	B	C	D	X^2	
pre	.36	.04	.05	.53	5.555	$.02 > p > .01$
post	.33	.15	.09	.44		



22. Are J and K similar triangles?

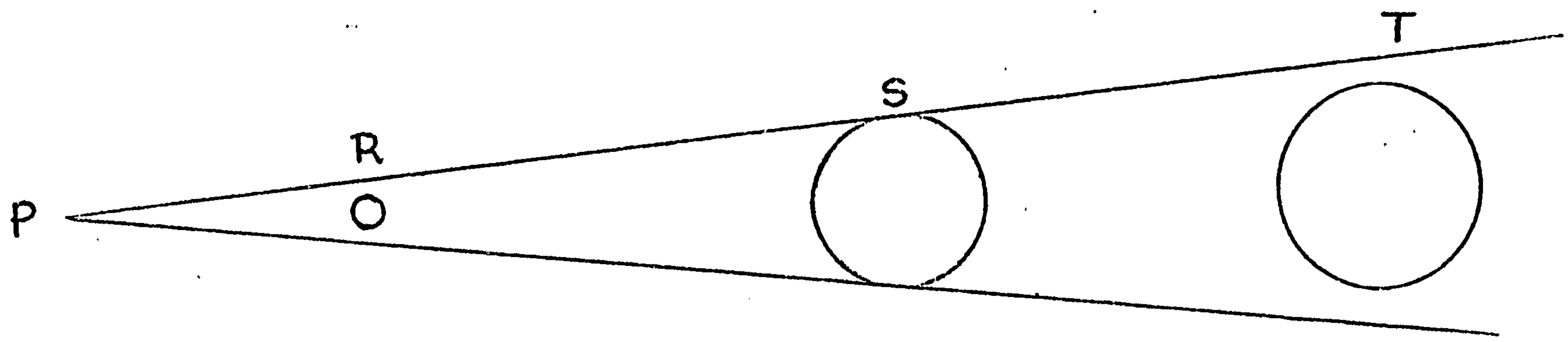
A. Yes. They look very similar.
B. Yes. The sides are almost the same size.
C. No. They are congruent.
*D. No. The angles in similar triangles must be the same.

	A	B	C	D	X^2	
pre	.14	.24	.17	.43	7.363	$.01 > p > .005$
post	.16	.09	.12	.63		

TABLE 3 (cont'd)

28. In the diagram below, which circle appears to be the largest when viewed from point P?

A. Circle R.
*B. Circle S.
C. Circle T.
D. They all appear to be the same size.
E. R and T appear larger than S.



	A	B	C	D	E	X^2	
pre	.15	.07	.55	.15	.02	8.894	$.005 > p > .001$
post	.23	.22	.42	.10	.01		

on both pre and posttest, and 69 had #28 wrong both times. While the statistic denotes significant shift, the results for those two items on the posttest do not at all indicate mastery.

In a further attempt to discern more precisely what pupils learned or did not learn in the course of time spent with the ESSP materials, the posttest results on Book 1 questions were examined for items mastered by fewer than 25% of the pupils. (See Table 4) Three such items were found; they are numbers 13, 18, and 25.

Question #13 involves taking the idea of a scale model one step further than the way it is presented in the ESSP materials. Thirty-three per cent of the pupils choose incorrect response E on the posttest, however, which indicates that the concept that a scale model involves a ratio at least was known by one third of the pupils.

Question #18 requires information learned in Chapter 7. As the time allotted for this study came to a close, this final chapter did not receive attention comparable to that for the other chapters. This may account for poor pupil performance on this item.

An explanation for the poor performance on #25 can be found in the very popular incorrect response C. Pupils apparently learned the applicable relationship: rate times time equals distance. Sound at the rate of 1200 ft. per sec. would travel 3600 feet in 3 seconds. But the problem discusses an echo and the pupils failed to recognize that an echo requires that sound travels to a certain point and then returns.

TABLE 4

Book 1 Items from the Pretest Mastered by Fewer than 25% of the Students on the Posttest

13. To find the scale of a model boat, you would

A. find the difference between the length of the model boat and the length of the real boat.
B. measure both the length of the mast and the length of the sail since at least two measurements are always needed.
* C. divide the length of the sail on the model boat by the length of the sail on the real boat.
D. multiply the length of the model boat by the length of the real boat.
E. divide the length of the real boat by the length of the model boat.

	A	B	C	D	E
pre	.28	.19	.08	.12	.27
post	.36	.10	.10	.09	.33

18. A lamp is 10 feet away. The light seems dim, so you move to a chair 5 feet away from the lamp. The brightness of the light is now

A. 5 times as much.
B. 10 times as much.
C. two times as much.
* D. four times as much.

	A	B	C	D
pre	.42	.03	.53	.01
post	.26	.02	.70	.01

25. A hunter firest his rifle near a cliff. He hears the echo of his shot 3 seconds later. How far away is the hunter from the cliff? (Sound travels at about 1,200 feet per second.)

* A. 1,800 feet.
B. 2,400 feet.
C. 3,600 feet.
D. 7,200 feet.

	A	B	C	D
pre	.10	.03	.76	.08
post	.14	.03	.78	.04

General Knowledge Items: Table 5

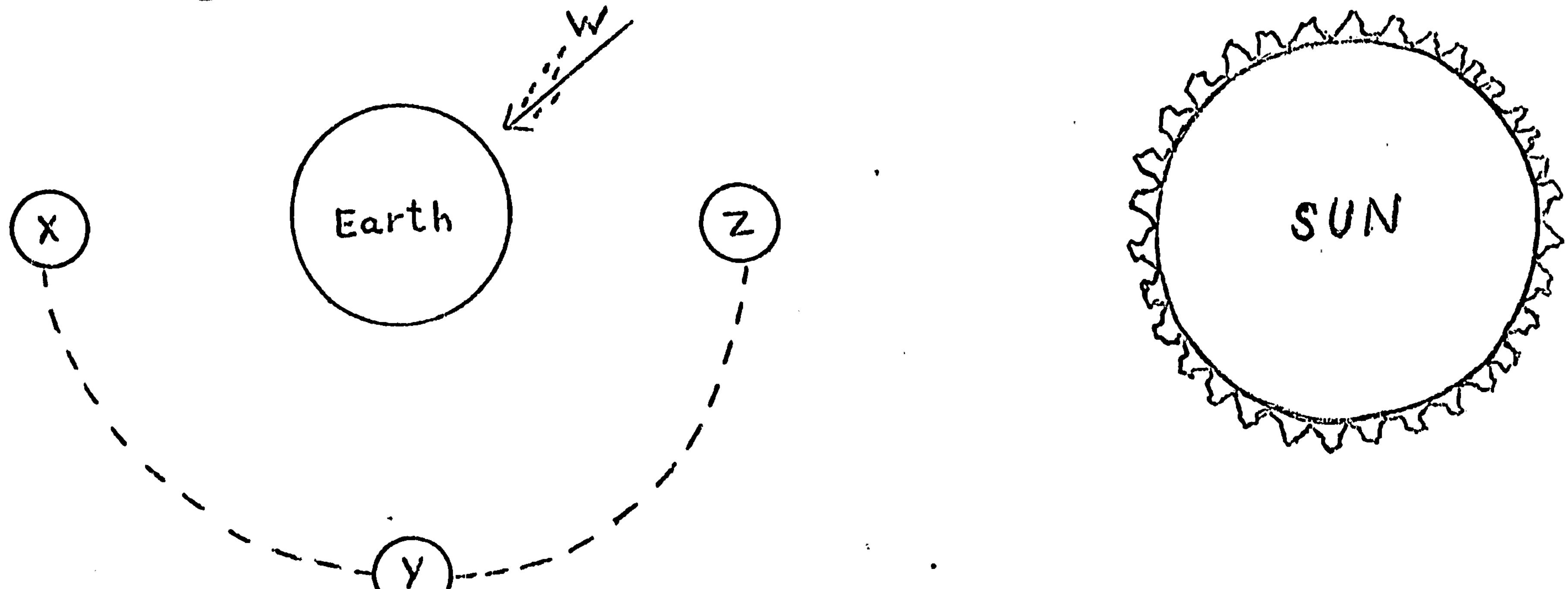
These items were questions about astronomy topics which were not part of the ESSP materials or, if they were present in the book, were incidental to what was being taught. The questions were included, as indicated above, to test the notion that this approach to the study of astronomy would also result in a concomittant increase in students' general knowledge of astronomy even though such information was not taught in Book 1. Nearly half (6 of 13) of the general knowledge items revealed a significant shift from pretest to posttest: #3, #5, #17, #20, #23, and #24. As can be seen by inspecting Table 5, the items cover a range of topics and only three could be considered to indicate mastery: #5 (76% of the pupils answered it correctly on the posttest), #17 (73% correct), and #23 (80% correct). And all of these three were answered correctly by more than half of the pupils on the pretest: 53 of 92 knew #5, 48 knew #17, and 55 knew #23.

The findings then indicate a significant trend toward increased knowledge of general information about astronomy, but do not reveal mastery of the information measured by the items from pre to posttest. It was not the intent of the ESSP materials to teach this general knowledge, so this aspect of assessment of these materials is not necessarily concerned with mastery. That a significant trend does exist showing an increase in the pupils' general knowledge of astronomy is the important finding.

TABLE 5

General Knowledge Items with Significant Chi Square Test of Change: Proportions of Responses by Students

Drawing for Questions 1 to 5.



3. When the moon is at position Z, there could be an eclipse of the

A. earth
* B. moon
C. sun

	A	B	C	X^2	
pre	.13	.23	.65	4.0000	p= .05
post	.14	.36	.52		

5. The small particles at W in the drawing are falling through the earth's atmosphere. They are probably

* A. meteorites
B. comets
C. stars
D. galaxies
E. asteroids

	A	B	C	D	E	X^2	
pre	.66	.09	.07	.05	.13	4.166	p= .05
post	.76	.14	.03	.03	.03		

TABLE 5 (cont'd)

17. Which of the following lists contains <u>only</u> planets?

* A. Neptune, Pluto, Uranus, Mercury.
B. Jupiter, Venus, Sputnik, Earth.
C. Earth, Mars, Moon, Jupiter.
D. Uranus, Saturn, Sun, Phobos.

	A	B	C	D	X^2	
pre	.56	.01	.33	.09	9.78	$.005 > p > .001$
post	.73	.01	.23	.03		

20. Which list of planets is in the correct order of increasing distance from the sun?

A. Mars, Earth, Venus, Jupiter.
* B. Venus, earth, Mars, Jupiter.
C. Earth, Mars, Jupiter, Venus.
D. Mars, Jupiter, Earth, Venus.

	A	B	C	D	X^2	
pre	.25	.40	.24	.11	9.322	$.005 > p > .001$
post	.20	.59	.11	.11		

23. Which of the following is the best description of <u>constellations?</u>

A. Wandering stars.
* B. Groupings of stars.
C. Non-moving stars.
D. Lines which connect stars.

	A	B	C	D	X^2	
pre	.05	.65	.05	.20	8.166	$p = .005$
post	.05	.80	.02	.10		

24. Which list of objects is in the order of increasing size?

A. Solar system, Sun, Moon, Earth.
B. Earth, Moon, Sun, Solar system.
*C. Moon, Earth, Sun, Solar system.
D. Sun, Moon, Earth, Solar system.

	A	B	C	D	X^2	
pre	.17	.12	.53	.14	5.761	$p = .025$
post	.13	.11	.65	.11		

Book 1 Items on Posttest only: Table 6

There were nine Book 1 items on the posttest which were not included on the pretest. They were only included on the posttest because it was believed that the pupils did not possess the knowledge and understanding required by these items prior to studying the ESSP materials. Pupil mastery was not achieved on any of the five items which required recall of factual knowledge: #29, #31, #32, #33, and #37. Two items in fact were answered correctly by fewer than 25% of the group (#32 and #37). This may be a reflection of the authors' intent to place "great weight on a few fundamental concepts of astronomy rather than on a scattering of isolated facts." (Teacher's Guide, page iii)

Three items, #30, #35, and #36, required the pupils to apply a rule or principle learned with Book 1 materials. While mastery was not achieved on any of these items, none fell below the 25% level. One item, #34, requires knowledge of specific facts (Eratosthenes experiment) but also requires understanding of the underlying principles which organize the facts in order to apply then to a new situation. Pupil performance on this item approaches mastery (67% correct).

Grouping these nine items into three classes: recall of facts, application of principles, and knowledge of facts and organizing principles, from a lower to higher order of cognitive process involved, there also emerges a lower to higher pattern of pupil performance. Inasmuch as the authors' purpose was to

TABLE 6

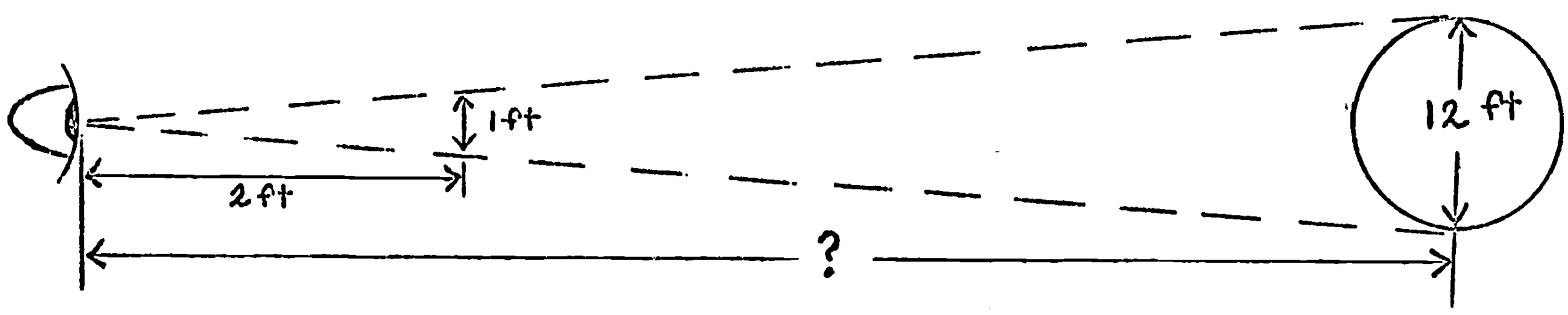
Book 1 Items Which Were on the Posttest Only: Proportions of Responses by Students

29. As you move an object away, it appears to become smaller because the

A. apparent angle increases.
* B. angular size decreases.
C. atmosphere is hazy.
D. earth's surface is curved.
E. light has to travel further.

A	B	C	D	E
.21	.47	.02	.19	.12

30. You launch a balloon which is 12 feet in diameter and watch it rise. You then hold up a one foot ruler 2 feet in front of you. The ruler just covers the balloon from one edge to the other edge. How far away is the balloon?



A. 6 feet
B. 12 feet
* C. 24 feet
D. 30 feet
E. 48 feet

A	B	C	D	E
.44	.21	.25	.05	.03

31. When Kepler made up the name Astronomical Unit it stood for

A. the diameter of the earths orbit.
B. 186,000,000 miles.
C. the distance the earth travels in one year.
D. the length of time it takes light to travel from the sun to the earth.
* E. the distance from the sun to the earth.

A	B	C	D	E
.19	.12	.10	.05	.53

TABLE 6 (cont'd)

32. To find the distance to the sun astronomers did not use the principle of the range finder because

* * A. the angles are too difficult to measure.
* B. the sun is too big.
* C. it is impossible to direct a range finder at the sun.
* D. the earth moves around the sun.

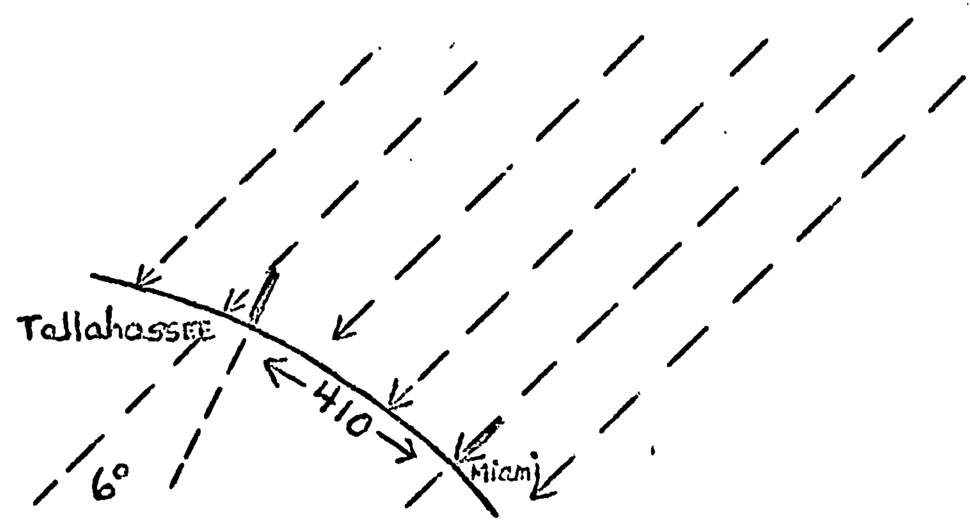
A	B	C	D
.23	.10	.30	.36

33. To locate the orbit of an inner planet correctly on a scale model of the solar system, astronomers need to find the planet's

* A. distance from the earth.
* *B. maximum angular separation from the sun.
* C. maximum distance from the sun.
* D. size and average length of day.

A	B	C	D
.20	.49	.20	.10

Tom lives in Miami, Florida, and Jerry lives in Tallahassee. One sunny afternoon Tom noticed that the telephone poles cast no shadows. He ran into the house and called up Jerry. Tom asked Jerry to measure the angle of the shadow of a telephone pole near his house. Jerry told Tom that the angle was 6 degrees. Tom looked at a map and found that the distance between Miami and Tallahassee is 410 miles.



34. Tom wants to repeat Eratosthenes' calculation for getting the circumference of the earth. Can Tom make the calculation with the information he has?

* A. No. The calculation works only for Aswan and Alexandria.
* * B. Yes. The calculation works anywhere on the earth.
* C. No. The calculation was good for the ancient Greeks but is not practical nowadays.
* D. Yes. If Tom also finds out the distance from Miami to Aswan.

TABLE 6 (cont'd)

E. No. One of the angles has to be measured in a well.

A	B	C	D	E
.04	.67	.10	.03	.15

35. If Tom did make Eratosthenes' calculation with his data, what value would he get for the circumference of the earth?

A. 2,460 miles
B. 12,500 miles
* C. 24,600 miles
D. 25,000 miles
E. 41,000 miles

A	B	C	D	E
.16	.10	.39	.17	.12

36. If the diameter of a circle is 2 feet, its circumference is about

A. π feet
B. 1.5 feet
C. 3 feet
D. 3.14 feet
* E. 6 feet

A	B	C	D	E
.14	.09	.10	.15	.50

37. We divide a circle into 360 degrees because

A. a straight line must be a 180 degree angle.
* B. people agreed to do it that way.
C. an early Greek measured and found there were 360.
D. there is one degree for each section of the sky.
E. 360 is a universal constant.

A	B	C	D	E
.48	.12	.13	.12	.13

focus on larger concepts rather than on isolated facts, the pupils' performance reflects success with the material in the direction intended by those who designed the materials.

Skills Items: Table 7

The last five items on the posttest were designed to test whether the pupils had acquired certain skills in making measurements, skills used by astronomers in learning about the universe.

Each item was scored by two judges working independently and using a four point scale: 0, 1, 2, and 3. Pupils who did not even attempt an item were given 0, so that in effect a 0 score indicates an omission. A score of 3 indicates a complete and precise measurement. The other ranks on the scale, 1 and 2, were given to answers which were incomplete or showed less precise measurements than the criterion set for a rating of 3. The amount of error acceptable for each rating was decided upon beforehand, and the two raters agreed in all but a few cases which were then reviewed by the raters together.

Items #38, #39, and #42 required the pupils to make some simple measurements using a ruler, protractor, and ruler and protractor respectively. Many pupils, 83 to 90%, got a rating of 2 or 3 on these items. (See Table 8) The pupils in this study seem to be able to use these tools effectively. Pupil performance on items #40 and #41, however, was not comparable, those being mastered by only 43% and 26% of the population. These two items required computation beyond the simple measurement, and this additional skill was apparently not learned by the pupils from using the ESSP materials.

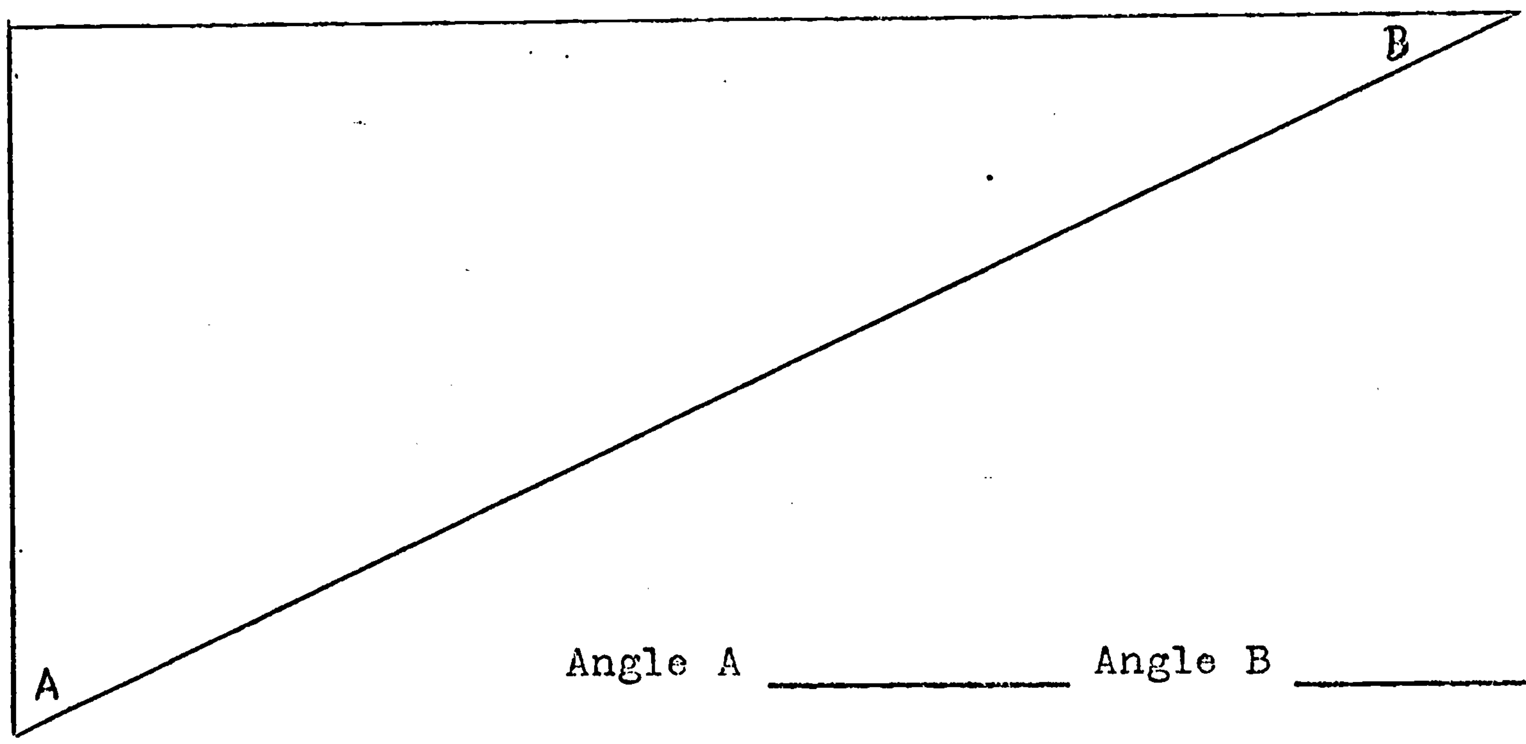
TABLE 7

Skills Items

38. Measure the length of the line segment below with your ruler. Write the best value for the length in the answer box.

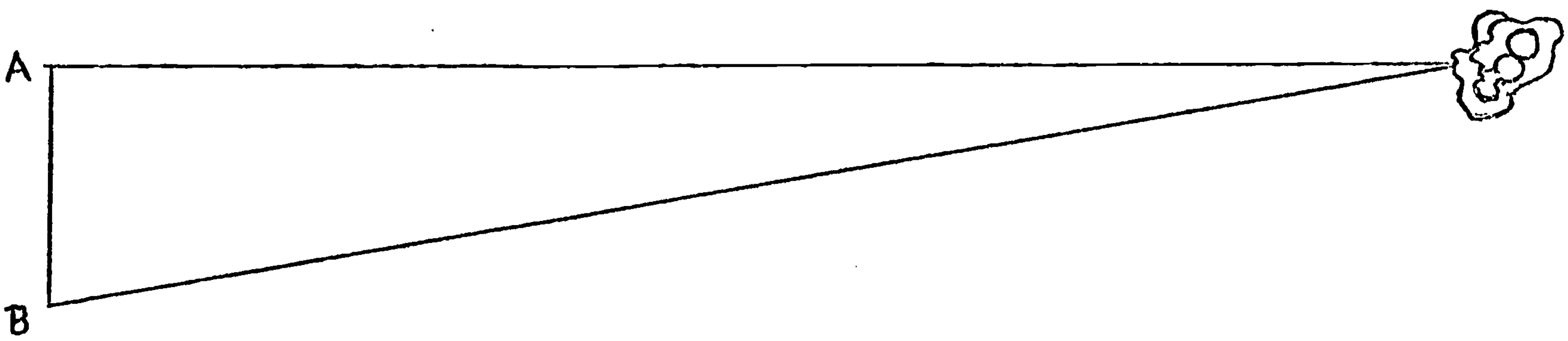
Answer

39. Measure angle A and angle B with your protractor. Write the values in the answer spaces.



Angle A ____________ Angle B ______________

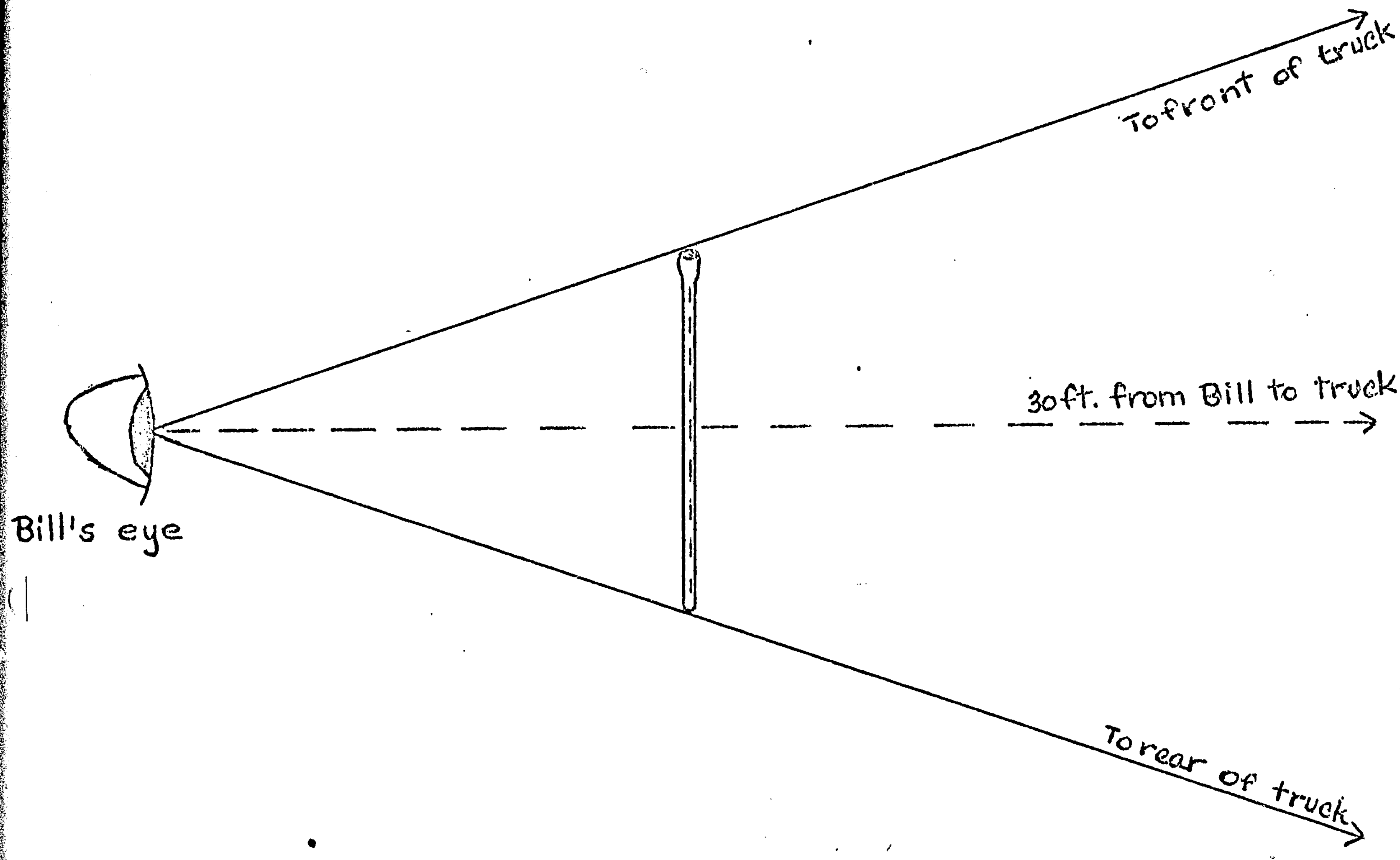
40. The drawing below is a scale drawing made with a range finder to find the distance to the tree. The base line of the range finder A-B is 1.5 feet long. Measure with your ruler and then calculate the actual distance to the tree. Record your answer in the space provided.



Distance to tree ____________

TABLE 7 (cont'd)

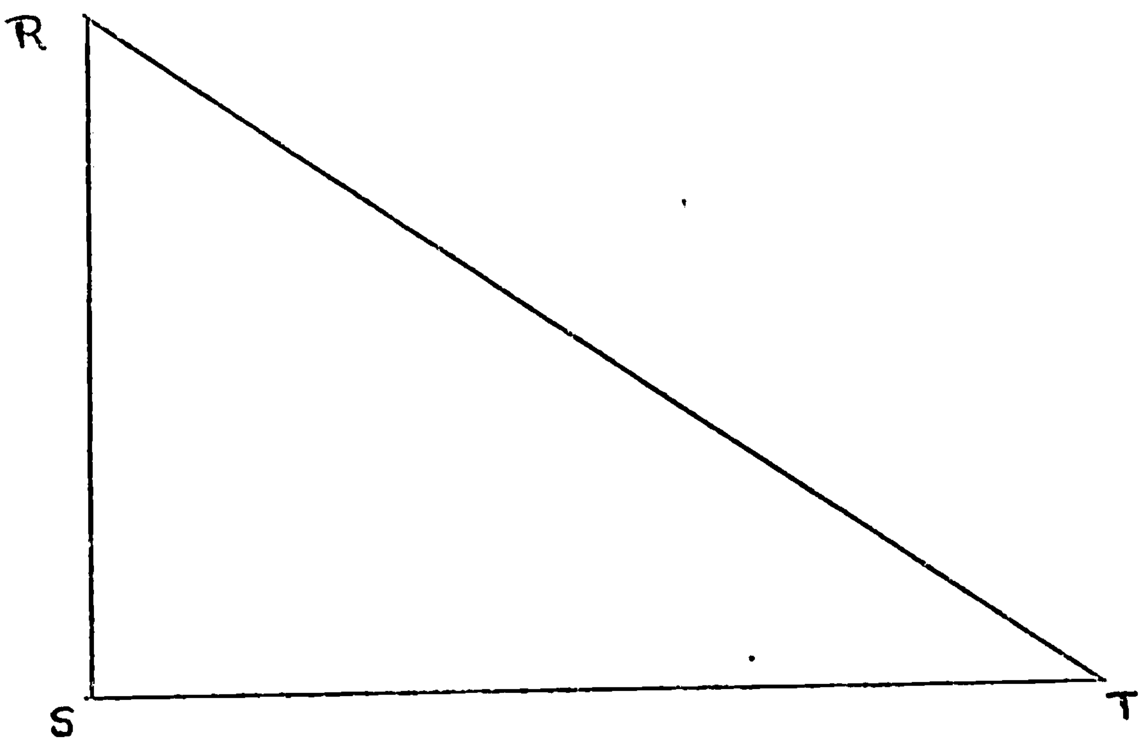
41. Bill wanted to find the length of a truck which was 30 feet away. He held a matchstick near his eye so that it just covered the truck from front to rear. Below is a <u>full size</u> drawing of what Bill saw. Make the correct measurements and calculate the length of the truck. Write your answer in the space provided.



Length of truck ____________

TABLE 7 (cont'd)

42. Using the line segment L-M as one side, draw a triangle RST. Use your ruler and protractor.



L M

TABLE 8: Proportion of Students Rated at Each Level for Skills Items (N = 92)

ITEM #	omit	1	2	3	Mean * Score
38	0.	.03	.36	.61	2.58 (92)
39	0.	.10	.20	.71	2.61 (92)
40	.07	.51	.09	.34	1.81 (86)
41	.21	.54	.07	.19	1.55 (73)
42	.04	.12	.26	.58	2.48 (88)

* Computed on the basis of pupils who gave some response. Number of cases in parentheses.

B. EFFECTS ON STUDENTS OF THE ESP MATERIALS

Test On Understanding Science

The third hypothesis proposed concerned an increase in "students' general understanding of science (as measured by the Test On Understanding Science)." To test this hypothesis, a *t* for correlated groups was computed from the scores of the TOUS pretest and posttest. (See Table 9.)

The computed value of *t* indicates a significant difference at the .05 level between TOUS pre and post test means. But the actual difference between the means is only .88, a mean change of less than one item correct from pre to posttest. Closer inspection of the data seemed a possible source of more meaningful information about what happened from pre to posttest administration and about which items contributed most to this change in mean score.

Again, to employ the McNemar chi square test of change, contingency tables were formed for the 36 items on the TOUS test. Those items which revealed a significant shift are shown in Table 10.

Seventy percent of the pupils having chosen the "best" answer was used as an index here, parallel to the 70% level of mastery used in the analyses of the subject-matter achievement test items. Of the six items with significant chi squares, on only three was the best answer chosen by 70% or more of the pupils on the posttest. (#5, #12, and #26.) And it is of interest to note that these three "best" answers were known by more than 50% of the pupils on the pretest, the best answer to #12 even having been chosen by almost 70%.

TABLE 9

Comparison of Pretest and Posttest Scores on TOUS
(N = 89)

	Mean	S.D.	t	
Pretest	19.92	4.111	2.01	p = .05
Posttest	20.80	4.399		

TABLE 10

TOUS Items with Significant Chi Square Test of Change: Proportions of Responses by Students

5. Which of the following sentences about science is best?

A. Modern science is too advanced to use past discoveries.
B. Modern science develops modern products.
C. Modern science depends on useful inventions.
* D. Modern science is based on the science of the past.

	A	B	C	D	X^2	
pre	.07	.26	.14	.52	9.142	$.005 > p > .001$
post	.03	.16	.11	.70		

6. A scientific theory should

A. provide the final solution to scientific problems.
B. suggest directions for making useful things.
* C. tie together and explain many natural events.
D. suggest good rules for carrying out experiments.

	A	B	C	D	X^2	
pre	.23	.20	.38	.18	4.000	$p= .05$
post	.13	.20	.50	.17		

12. The scientists of today can work on more complex problems than the scientists of the past mainly because they

A. work harder than earlier scientists.
B. have more ideas than earlier scientists.
* C. build on the work of earlier scientists.
D. are more clever than earlier scientists.

	A	B	C	D	X^2	
pre	.01	.25	.69	.05	4.166	$p= .05$
post	.01	.15	.79	.04		

TABLE 10 (cont'd)

17. Which of the following is the main need of science?

* A. People with new ideas.
  B. More money and equipment.
  C. Well-trained craftsmen.
  D. Better working conditions.

	A	B	C	D	X^2	
pre	.52	.16	.19	.12	5.444	p= .02
post	.36	.27	.21	.16		

24. When a scientist makes a new discovery, he usually makes a report of it because he

  A. hopes to help mankind by announding his discovery.
  B. wants to prevent other scientists from making the same discovery.
* C. wants other scientists to know about his work and check it.
  D. hopes other scientists will help him to finish his work.

	A	B	C	D	X^2	
pre	.66	0.	.26	.09	16.030	$p > .001$
post	.34	.03	.52	.11		

26. Before a scientist announces a new theory to the public, he will most likely talk his ideas over with

  A. government leaders who many want to use his theory.
* B. other scientists in his special field.
  C. science writers of large newspapers.
  D. a group of experts on scientific theories.

	A	B	C	D	X^2	
pre	.13	.59	0.	.27	8.333	$p > .005$
post	.08	.77	.04	.11		

Two of the remaining three items, #6 and #24, were known by approximately half of the pupils on the posttest (50% and 52% respectively). And the last of the answers which showed a significant chi square for change (#17) shifted in the negative direction. Fifty-two percent of the pupils chose the best answer on the pretest, while only 36% chose it on the posttest.

A second kind of closer look at TOUS test results was an examination of individual students' pretest to posttest changes. Of the 89 students for whom complete pre and posttest data on the TOUS are available, 37.1% achieved lower scores on the posttest than they had on the pretest. This drop ranged from -1 to -9, the mean loss being -3.42. In 53 cases there was an increase from 1 to 13 points, the mean gain being 3.62, and three pupils' TOUS scores were the same on both the pretest and posttest.

From the preceding evidence, it is difficult to find support for our hypothesis that study of ESSP materials will "increase students' general understanding of science (as measured by the Test On Understanding Science)." The t test indicates a statistically significant (.05 level) difference between the pretest mean and the posttest mean, but on very few posttest items (3) did 70% or more of the pupils choose the best answer, and a large percentage of the population (37.1%) appeared to have decreased "general understanding of science" as measured by this instrument.

Our fourth hypothesis concerned "general understanding of science" also. It posited that "a student's gain in general understanding of science is positively correlated with subject matter achievement." To test this idea, two coefficients were computed to examine the correlation between the pre to posttest differences on the subject-matter test with the pre to posttest differences on TOUS. As can be seen from Table 11, two "difference scores" were calculated for the subject-matter tests. One was the pre to posttest difference on Book 1 items, the 15 questions which appeared on both tests. (This difference, therefore, does not include the 9 Book 1 items which appeared on the posttest only.) The other subject-matter pre to posttest difference which was calculated was the difference between the total pretest score (28 items including Book 1 and General Knowledge questions) and the total posttest score (42 items, consisting of the 28 pretest items, an additional 9 Book 1 items, and 5 skills items). In this way we hoped to see whether gains in achievement on ESSP materials was correlated with increase in general understanding of science, and also whether the results of the total experience, as measured by the posttest, with the achievement on the pretest subtracted out, would be related to gains on TOUS. In the table the former difference is labelled "Book 1 difference" and the latter is "Total Test difference."

While our earlier discussion indicates that little evidence can be found for real differences in TOUS scores from pre to posttest administration, we were still interested to know

TABLE 11

Correlations Between Pretest to Posttest Differences

	TOUS Difference = .876	
	Correlation	Confidence Interval
Book 1 Difference = 1.876	.386	.95 (.196 ζ .618)
Total Test Difference = 8.685	.198	.95 (.010 ζ .412)

whether what change we did find was at all related to subject-matter achievement. The indices generated by our correlation of test score differences does not seem to support the hypothesis. Quite the contrary, it seems almost a chance occurence that more pupils gained than lost from pre to posttest on the TOUS.

Word Association Study

The semantic differential instrument, "Word Association Study" (WAS), was designed to test our fifth hypothesis, viz., "Study of ESSP materials will affect students' views of astronomy, arithmetic, scientists, and learning experiences in science." The WAS instrument, as previously described in this report, yielded pupils' ratings of eight concepts on each of 15 bipolar adjectival scales. Thus, the data collected for testing our fifth hypothesis consisted of 120 pairs of ratings from the pre and posttest administration of the WAS instrument.

(In the WAS instrument, ratings of 1 to 5 were assigned to the five positions on each scale. This procedure assumes an equality of the intervals between scale positions. Although this assumption was not tested in this study, there is support for it from previous research with the semantic differential, and our procedure of assigning equal interval ratings is commonly used. Furthermore, some of the scales on the WAS instrument are "reversed" as printed in the test booklet to counteract a possible response set on the part of the student. The ratings of these "reversed" scales are converted during the analysis so that the numerical values of scales with similar meanings will be consistent. Thus, for example, the fifth scale as printed on the WAS instrument is "important-unimportant", but this "reversed" scale is converted during the analysis into an "unimportant-important" scale with a rating of 1 assigned to the "unimportant" pole and ratings of 2, 3, 4, and 5 assigned to the other scale positions so that the order is consistent with increasing "importance.")

The means and standard deviations of the pretest and posttest ratings were calculated for the 120 items on the WAS instrument, and a representative selection of these is presented in Table 12. The table also shows the changes in mean ratings of the several scales from pre to posttest, and we may take note of a number of these shifts that are rather suggestive. In commenting on these shifts, we are aware that only limited confidence can properly be placed in the ratings for any one scale on a semantic differential instrument; hence we treat the changes in mean ratings as suggestive, rather than as strong evidence of changes in students' perceptions. We estimate that, generally, a change in mean rating for an item in excess of 0.20 is statistically significant at the .05 level, and that a change in excess of 0.26 is statistically significant at the .01 level. (This estimate is based on calculations of a <u>t</u> for correlated means for a random sample of ten items in the table.) On this basis, then, we note in the data presented in Table 12 that:

Students seemed to view ASTRONOMY as less powerful, less exciting, and less enjoyable on the posttest than on the pretest. From pre to posttest, they came to regard it as less difficult,

TABLE 12

Pretest to Posttest Changes in Mean Ratings of Selected Concepts and Scales on the Word Association Study (N=92)
(Minimum rating = 1; maximum = 5)

Concepts and Scales	Pretest Mean	Pretest S.D.	Posttest Mean	Posttest S.D.	Pretest to Posttest Change in Mean
ASTRONOMY					
weak-powerful	4.054	0.864	3.739	0.924	-0.315
hard-soft	2.413	0.963	2.565	0.856	+0.152
unimportant-important	4.828	0.524	4.587	0.854	-0.241
dull-exciting	4.602	0.809	3.815	1.240	-0.787
mannish-womanish	2.359	0.897	2.522	0.733	+0.163
unenjoyable-enjoyable	4.355	0.917	3.924	1.188	-0.431
changing-permanent	1.796	1.119	1.804	1.008	+0.008
easy-difficult	3.806	0.912	3.543	0.907	-0.263
ARITHMETIC					
weak-powerful	4.000	0.921	4.011	1.093	+0.011
hard-soft	2.473	0.951	2.700	0.893	+0.227
unimportant-important	4.957	0.204	4.856	0.628	-0.101
dull-exciting	3.710	1.247	3.856	1.286	+0.146
mannish-womanish	2.925	0.448	2.911	0.630	-0.014
unenjoyable-enjoyable	3.860	1.166	4.022	1.281	+0.162
changing-permanent	2.355	1.486	2.644	1.352	+0.289
easy-difficult	3.247	1.204	2.911	1.088	-0.336
READING ABOUT SCIENCE					
slow-quick	3.000	1.063	3.130	1.121	+0.130
hard-soft	3.032	0.865	2.890	0.781	-0.142
unimportant-important	4.796	0.523	4.620	0.850	-0.176
unenjoyable-enjoyable	4.152	1.026	3.783	1.184	-0.369
useless-useful	4.753	0.564	4.511	0.978	-0.242
easy-difficult	2.763	1.026	2.652	0.999	-0.111
MAKING MEASUREMENTS					
slow-quick	3.054	1.087	3.043	1.128	-0.011
hard-soft	2.828	0.880	2.707	0.846	-0.121
unimportant-important	4.591	0.741	4.641	0.673	+0.050
unenjoyable-enjoyable	3.828	1.185	3.761	1.152	-0.067
useless-useful	4.667	0.727	4.565	0.789	-0.102
easy-difficult	2.968	0.983	2.902	0.973	-0.066
MOST SCIENTISTS					
weak-powerful	4.097	0.873	3.891	0.931	-0.206
dirty-clean	4.183	1.032	3.891	1.104	-0.292
dull-exciting	4.376	0.932	4.065	1.003	-0.311
mannish-womanish	2.409	0.837	2.489	0.832	-0.080
moving-still	1.914	0.974	2.141	1.044	+0.227
useless-useful	4.871	0.396	4.652	0.907	-0.219
foolish-wise	4.710	0.600	4.641	0.793	-0.069
boring-interesting	4.527	0.788	4.087	1.096	-0.440

TABLE 12 (cont'd)

Concepts and Scales	Pretest Mean	Pretest S.D.	Posttest Mean	Posttest S.D.	Pretest to Posttest Change in Mean
MY SCIENCE TEACHER					
weak-powerful	3.871	0.923	3.253	1.131	-0.618
dirty-clean	4.495	0.916	4.000	1.211	-0.495
dull-exciting	4.301	0.906	3.692	1.347	-0.609
mannish-womanish	4.559	0.853	4.527	1.015	-0.032
moving-still	1.624	0.793	2.297	1.206	+0.673
useless-useful	4.699	0.749	4.407	1.164	-0.292
foolish-wise	4.667	0.771	4.187	1.264	-0.480
boring-interesting	4.462	0.854	3.802	1.368	-0.660

and there was no shift in their view of astronomy as changing or permanent.

Between pretest and posttest, the pupils shifted in their view of ARITHMETIC toward seeing it as less difficult. It also seemed to become "softer" for them and more permanent.

READING ABOUT SCIENCE was viewed as less enjoyable by the students on the posttest than on the pretest.

The pupils' view of MOST SCIENTISTS seemed to shift from pre to posttest toward less clean, less exciting, and less interesting. The shifts in the view of MOST SCIENTISTS are fairly closely matched, though to a less degree on each scale, with the shifts in the view of MY SCIENCE TEACHER.

The SCIENCE TEACHER appears to have suffered some losses in the pupils' view from pretest to posttest. On the second occasion, she was seen as less powerful, less clean, less exciting, less useful, less wise, less interesting, and more still.

While changes such as these in mean ratings on single scales for particular concepts are suggestive, further analysis of the responses to the WAS instrument provided information of somewhat greater interest about the pupils' perceptions. The response data generated from the administration of a semantic differential instrument lend themselves admirably to factor analytical techniques for the purpose of uncovering the underlying structure represented by the responses. The aim of the factor analysis is to discern common factors made up of scales on the semantic differential for which pupils' responses tend to cluster together. To use Osgood's terminology, these factors represent dimensions of the pupils' meaning space. We prefer to think of these factors as constitutent elements of the pupils' images of particular concepts. The elements of the image of a concept may be either cognitive or attitudinal. The nature of the scales selected for a given semantic differential instrument determines whether cognitive elements, attitudinal elements, or both, may be discerned in the analysis of the responses on the instrument. For the WAS instrument, we selected adjectival scales that were primarily attitudinal, so that we would expect the factor analyses to reveal some of the attitudinal elements of the pupils' images of the concepts.

The computer program which we used performed a principal components factor analysis of the intercorrelation matrix and, when called for, rotated factor analysis was made of the 15 scales of the WAS instrument across all eight concepts. This analysis produced one (unrotated) factor that accounted for 65.8% of the variance. Scales with high factor loadings on this factor included: unimportant-important, dull-exciting, bad-good, unenjoyable-enjoyable, useless-useful, foolish-wise, boring-interesting. Further preliminary analyses showed evidence of considerable interaction between scales and particular concepts, indicating that there was no single factor structure common to all the concepts rated on the WAS instrument. Hence, we subsequently factor analyzed separately each concept (or page) of the WAS instrument across the 15 scales on which the concept was rated. These analyses showed that there were three WAS instrument concepts for which the factor structure was very nearly the same.

Table 13 presents three principal (rotated) factors that our factor analyses revealed for the concepts ASTRONOMY, MAKING MEASUREMENTS, and DOING SCIENCE EXPERIMENTS. The cumulative

TABLE 13

Three Principal Factors and Percent of Variance for Three Concepts on the Word Association Study

CONCEPT:		ASTRONOMY	MAKING MEASUREMENTS	DOING SCIENCE EXPERIMENTS
		Percent of Variance (5 rotated factors)		
Factor I	PERSONAL ENJOYMENT dull-exciting bad-good unenjoyable-enjoyable boring-interesting	17.2	19.0	23.0
Factor II	IMPORTANCE unimportant-important useless-useful foolish-wise	13.0	12.2	16.8
Factor III	DYNAMISM slow-quick weak-powerful	12.1	9.6	13.1

percentage of the total variance accounted for by these three factors typically ranges between 40% and 52%. (The percentages shown in the table are based on pretest data for ASTRONOMY and MAKING MEASUREMENTS, and on posttest data for DOING SCIENCE EXPERIMENTS.) The table also shows the scales that had high factor loadings on each of the three factors, and it was from the meanings associated with the included scales that we constructed the names assigned to each factor. Factor I includes scales that seem to represent a child's personal involvement; the clue here, we believe, is that a fifth-grade child generally talks of something as either "exciting" or "bad" or "good" in terms of his own experiences, rather than in an abstract sense. Hence, we gave the name "Personal Enjoyment" to Factor I. The scales included under Factor II, on the other hand, do not seem to reflect this same personal involvement, but point to a more detached and somewhat more sophisticated evaluation of concepts on the part of these fifth graders. In fact, Factor II appears to have a close kinship with the "Evaluative" factor repeatedly found in semantic differential studies with adults as subjects. We chose the name "Importance" for Factor II. Consideration of the two scales included under Factor III suggested that "Dynamism" would be an appropriately descriptive name for this factor. We have, then, three constituent elements of the pupils' images of three concepts which they rated on the WAS instrument, elements that denote the personal enjoyment, importance, and dynamism of these concepts as the pupils viewed them.

What were the changes, if any, in the pupils' views of those concepts during the time that they studied the ESSP materials? To answer this question, we computed a composite score for each principal factor identified for each of the concepts, ASTRONOMY, MAKING MEASUREMENTS, and DOING SCIENCE EXPERIMENTS, from both the pre and posttest data on the WAS instrument. The composite score for a factor that includes k scales is the sum of 1/k times the rating of each scale. Thus, for example, the composite score for "Importance" is equal to the sum of 1/3 times the rating on the "unimportant-important" scale, 1/3 times the rating on the "useless-useful" scale, plus 1/3 times the rating on the "foolish-wise" scale. We also calculated the pre and posttest means and standard errors of the composite scores for "Personal Enjoyment," "Importance," and "Dynamism" on each concept and, using a t-test for correlated data, made comparisons of the nine pretest-posttest means. (See Table 14.)

As is displayed in the table, the students' image of ASTRONOMY decreased in the element of "Personal Enjoyment" (change in mean composite score significant at the .01 level) and in the element of "Dynamism" (change in mean composite score also significant at the .01 level) during the time they were studying the ESSP materials. With reference to their personal enjoyment of astronomy, the pretest mean composite score for the group was extremely high (about 4.5 out of a possible maximum of 5), and the posttest mean composite score (about 4.0) was still very high. There was, however, a considerable spread-

TABLE 14

Pretest-Posttest Comparisons of Three Factor Means for Three Concepts on the Word Association Study

Factor		ASTRONOMY		MAKING MEASUREMENTS		DOING SCIENCE EXPERIMENTS	
		Mean	S.E.	Mean	S.E.	Mean	S.E.
I - Enjoyment	-Pre	4.519	0.067	4.106	0.084	4.657	0.050
	-Post	3.976	0.110	3.965	0.102	4.481	0.073
Change		-0.543	0.106	-0.141	0.107	-0.176	0.072
		(t= 5.13**)		(t=1.32)		(t=2.44*)	
II. -Importance	-Pre	4.584	0.058	4.557	0.058	4.710	0.040
	-Post	4.473	0.077	4.448	0.074	4.624	0.064
Change		-0.111	0.073	-0.109	0.078	-0.086	0.062
		(t= 1.52)		(t= 1.40)		(t= 1.34)	
III. - Dynamism	-Pre	3.576	0.075	3.234	0.084	3.359	0.075
	-Post	3.272	0.062	3.261	0.080	3.408	0.075
Change		-0.304	0.080	+0.027	0.096	+0.049	0.090
		(t= 3.80**)		(t= 0.28)		(t= 0.54)	

* = Significant at .05 level

** = Significant at .01 level

ing out of the group's personal enjoyment ratings of astronomy, as is shown in the increase in the standard error of the mean composite score from 0.067 on the pretest to 0.110 on the posttest. With reference to the students' image of the dynamism of astronomy, there was a statistically significant shift from pre to posttest, as already noted, toward a less dynamic view of the subject. The posttest mean composite score of about 3.3 (between a possible range of 1 to 5) indicates that, after study of the ESSP materials, the pupils' view of astronomy included an element of only moderate dynamism. Lastly, the students' image of the element of "Importance" in their concept of astronomy did not change significantly from pre to posttest, and the means of the composite score on both occasions (about 4.6 and about 4.5) were extremely high.

Turning to the concept MAKING MEASUREMENTS, no significant changes were found in the students' view from pre to posttest. Referring to the element of personal enjoyment in their view of making measurements, both the pretest mean (about 4.1) and the posttest mean (about 4.0) of the composite score were very high. The importance of making measurements in science is certainly stressed in the ESSP materials, but this element in the students' image of measurement was already extremely high on the pretest (mean composite score of about 4.6), so that rather little group gain was possible. The slight decrease in the mean composite score for "Importance" on the posttest is not statistically, or educationally, significant. With reference to the third factor on this concept, the pupils' image of making measurements included an element of only moderate dynamism, judging by means for the composite scores of about 3.2 on both pre and posttest.

The element of "Importance" in the pupils' view of DOING SCIENCE EXPERIMENTS showed extremely high means of composite scores, both on the pretest (about 4.7) and on the posttest (about 4.6), and the slight decrease in the mean on the posttest is not significant. For the same concept, the decrease from pre to posttest in the mean composite score for the element of "Personal Enjoyment" is statistically significant at the .05 level. It may be that the pupils' use of the ESSP materials had some effect on this element of ther perception of doing science experiments, but we are not inclined to become wildly concerned about a loss of less than 0.2 in mean composite score. We note that the element of enjoyment in the pupils' view of doing science experiments remained extremely high throughout, with a mean composite score of about 4.7 and about 4.5 on the pre and posttest, respectively. Finally, DOING SCIENCE EXPERIMENTS was perceived not very dynamically by the students, and there was no significant change in this element of their view.

Let us now see what bearing the foregoing analyses and discussion have on our hypothesis for this part of the study (Hypothesis 5). We have presented findings which show that the pupils' view of astronomy was changed during the time that they studied the ESSP materials. (See Table 14 and Table 12.) We have also presented some suggestive evidence of changes in the pupils' views of arithmetic and of scientists. (Table 12). We have found little evidence in support of the idea that study of ESSP materials will affect students' views of learning experiences in science. In summary, our work with the WAS instrument provided partial support for the fifth hypothesis, that students' views of astronomy, arithmetic, scientists, and learning experiences in science would be affected through their study of the ESSP materials.

CONCLUDING REMARKS

This study has investigated the effectiveness of the ESSP materials in increasing students' knowledge of astronomy and of how astronomical information is obtained. We found that the fifth grade students in the University of Chicago Laboratory Schools who studied the ESSP materials were moderately successful in mastering some of the topics that were taught. We also found gains in the students' knowledge about certain astronomical topics that were not specifically taught in the ESSP Book 1 materials, and this was interpreted as an increase in the students' general knowledge of astronomy. Detailed analyses of subject-matter achievement test items were made to determine more precisely what the students did learn and did not learn in their study of the ESSP materials.

The study also investigated the effect on students of studying the ESSP materials. We found that the effect of studying these materials on the students' general understanding of science was slight. The extent of this effect was explored through analyses of items on the Test On Understanding Science. Through the use of a semantic differential instrument, we investigated further the ESSP materials' effect on students' perceptions. We found that study of the ESSP materials did affect the students' view of astronomy, probably affected the students' views of arithmetic and of scientists, but did not seem to affect the students' views of learning experiences in science.

Perhaps more important than the specific findings of this study, however, are the procedures of analysis and interpretation which are illustrated here. The specific findings of the study provide some valuable information about the effectiveness and effects of the ESSP materials, but this information should not be looked upon as a firm evaluative judgment of these materials unless and until the study is adequately replicated with other students and in other school settings. On the other hand, the procedures used in the study have wide applicability in the evaluation of innovations in curriculum materials. These procedures provide specific information about what knowledge and which ideas students mastered successfully, about where they failed to attain mastery, and about changes in students' perceptions of the subject. The developer of new curriculum materials can learn a great deal from his evaluation efforts if he will make conscientious application of the analytical and interpretative procedures which are illustrated by the present study.

ACKNOWLEDGEMENTS

The investigator wishes to record his indebtedness to the persons who assisted in conducting this study and in preparing this report. Grateful thanks are extended to Dr. J. Myron Atkin and staff members of the University of Illinois Elementary School Science Project, for supplying materials and for financial support; to Barbara Wehr for cheerfully undertaking the teaching of the ESSP materials; to the fifth grade Lab School students for their essential participation; and to my research assistants, Fred Geis, Jr., E. Lawrence Liss, and Mary E. McCullough, for invaluable help. The typing of this report was conscientiously and cheerfully performed by Miss Barbara Lee, whose fine work is greatly appreciated.

ERRATA

page 13, line 6   READ: using the normal distribution of the statistic $Z(N-4)^{\frac{1}{2}}$.

page 37, Table 10

for item no. 24, READ: $p < .001$

for item no. 26, READ: $p < .005$

page 46, line 10   READ: when called for, rotated factors according to the varimax criterion. Our first factor analysis was made of the 15