DOCUMENT RESUME

ED 303 362 SE 050 351

AUTHOR	Chomsky, Carol; And Others
TITLE	Doing Science: Constructing Scientific Theories as an Introduction to Scientific Method. Scientific Theory and Method Project (STAMPS). Technical Report.
INSTITUTION	Educational Technology Center, Cambridge, MA.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
REPORT NO	ETC-TR-85-23
PUB DATE	Nov 85
CONTRACT	NIE-400-83-0041
NOTE	93p.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Cognitive Processes; Cognitive Psychology; Computer Software; *Computer Uses in Education; *Hypothesis Testing; Learning Processes; Learning Strategies; Pilot Projects; *Science Instruction; Scientific Concepts; *Scientific Methodology; Secondary Education; *Secondary School Science

ABSTRACT

The Scientific Theory and Method Project (STAMPS) is concerned with students' understanding of two areas of science in secondary schools. One is understanding what a scientific theory is, and the other is understanding the methods used in conducting scientific research. In the science curriculum these topics are often described as the process of science as opposed to scientific facts. Students, along with learning about the content of the different scientific disciplines, need an emphasis on the scientific process as well. Two questions are considered concerning students engaging in various activities presented to them: (1) how do students think about these problems; and (2) what kinds of activities and materials are best suited for this kind of introduction to the scientific method? Three components of a unit dealing with these questions have been developed and piloted. The computer program "The King's Rule" was used to introduce hypothesis formation and testing. The program "The Scientific Method" was used to study experimental conditions and the effects of phenomena under investigation. The third unit experiments with natural language phenonema. The pilot program is presented along with excerpts and transcripts taken during trials of the program. (MVL)

DOING SCIENCE: CONSTRUCTING SCIENTIFIC THEORIES AS AN INTRODUCTION TO SCIENTIFIC METHOD

Technical Report

November 1985

Educational Technology Center
Harvard Graduate School of Education
337 Gutman Library Appian Way Cambridge MA02138

Scientific Theory and Method Project (STAMPS)

Doing Science: Constructing Scientific Theories as an Introduction to Scientific Method

Technical Report

November 1985

Prepared by:

Carol Chomsky

Maya Honda

Wayne O'Neil

Chris Unger

Group members:

Diane Bemis

Carol Chomsky

Mary Greaves

Maya Honda

Robert E. Kilburn

Wayne O'Neil

Maxine Rosenberg

Chris Unger

Al Weinstein

Preparation of this report was supported in part by the National Institute of Education (contract #NIE 400-83-0041). Opinions expressed herein are not necessarily shared by NIE and do not represent Institute policy.

Acknowledgements

The members of STAMPS would like to thank the students who participated in piloting the curricular Unit materials at Cambridge Rindge-Latin School, Watertown High School, and Day Junior High in Newton.

We also want to thank Wayne Williams of Cygnus Software for allowing us to revise The Scientific Method software for our piloting.

Finally, we thank Max Katz, Judah Schwartz, Charles Thompson and Stone Wiske for their helpful comments on an earlier version of this report.

Contents

INTRODUCTION

The Scientific Theory and Method Project (STAMPS) is concerned with students' understanding of two areas of science in secondary schools. One is understanding what a scientific theory is, and the other is understanding the methods used in conducting scientific research. In the science curriculum these topics are often described as the "process" of science as opposed to scientific "facts". We recognize that students, along with learning about the content of the different scientific disciplines, need an emphasis on the scientific process as well.

The STAMPS group is examining students' ability to understand and to master various concepts within the domains of scientific theory and research methods. We would like students to come to appreciate that the goals of science include deriving principles and building theoretical constructs which account for observed facts. We would like them to understand how theories are developed and how they differ from and relate to observations, and why there is a need for and interest in constructing theories that explain and go beyond what can be observed. They should understand the role and purpose of experimentation in examining particular phenomena. It is important also for students to realize that a theory is a product of the scientist's mind, not of the data in the world. Such an understanding is important for the training of research scientists, and for a general lay understanding of scientific information as well.

Our approach is to involve students directly in problem-solving and eventually in theory construction by presenting them with problems which challenge already existing concepts, as well as their notions of what a scientific question is. We give them a problem--a phenomenon to be explained and facts to be accounted for--in an area where the explanation is conceptually available to the students. Conceptual accessiblity is crucial; if students have some intuitions about a phenomenon, the methods of science will help them explore these intuitions and, hopefullly, come to a clearer understanding of the phenomenon. We help them organize the data, develop descriptions of the data and hypothesize about the underlying principles involved. Through experimentation the students test their hypotheses and revise them as necessary, until they are satisfied that the resulting hypothesis adequately accounts for the observed data.

We are particularly interested in students' ability to form hypotheses and test them. We expect to find differences in students' ability to organize data appropriately so that they can form a reasonable hypothesis; to recognize what a counterexample is and to search for them; to revise a hypothesis upon finding counterexamples; to recognize when they have examined enough alternatives to conclude that they have indeed formulated a satisfactory and robust operating principle; and to understand the generalizability and explanatory power of the principle they derive.

As we observe students engaging in the various activities that we present to them, we have two questions in mind. The first concerns the students. How do they think about these problems? Do they, for example, exhibit competency levels with regard to the above aspects of hypothesis formation and testing, as a component of theory construction more generally? The second concerns the activities we give them to work with. What kinds of activities and materials are best suited for this kind of introduction to scientific method?

This year's work of the STAMPS group has provided useful information on both counts. We piloted a variety of materials for teaching scientific method with three groups of secondary school students. What we learned about the students and about the materials is the content of this report.

Content of the Unit

To address the goals outlined above, we have designated the following content to be included in the experimental Unit that we are developing.

Terms: experiment
- hypothesis
- counterexample
- inference
- theory
- variable(s): independent, dependent, controlled
- observation
- data

Concepts: explanation: how a theory accounts for data
- hypothesis formation
- prediction
- hypothesis testing: confirming (supporting) evidence, falsify a hypothesis, revise a hypothesis
- how to view experiments: differences between real experiments and computer simulations, idealization, experimental error

data vs. theory

the logic of problem solving

cause and effect vs. correlation

Activities: pose an experimental question

choose variables (independent, dependent, controlled)

set up an experiment

gather data and set up data tables

construct a graph

communicate results to others

Components of the Unit

Instructional materials prepared so far fall into three areas. A fourth area, still to be developed, is a laboratory science component. Here is a brief description of the three components of the Unit that have been developed and piloted.

1. Hypothesis formation and testing: computer-based "puzzle-solving" of non-natural phenomena.

   We chose the program, The King's Rule: Mathematics and Discovery (referred to here as simply The King's Rule; Sunburst Communications), to introduce hypothesis formation and testing. This program presents a number triad and the students' task is to formulate the rule which generates the triad. In doing so, the students cover the following content:

   terms: hypothesis, trial, counterexample

   concepts: forming, testing, falsifying and revising a hypothesis

   data may be consistent with more than one hypothesis

   account for all previous data when revising a hypothesis

2. Experimental method: microcomputer simulations of natural phenomena.

   We chose the program, The Scientific Method (Cygnus Software), to give students the opportunity to work with a variety of experimental conditions and observe the effects on the phenomenon under investigation. One example is an experiment on the chirp rate of crickets. This simulation allows students to manipulate a number of variables and, by working through experiments, measure the effect of the variables on the chirp rate. Students cover the following content:

   terms: experiment, hypothesis, counterexample, variables, observation, data

   concepts: the logic of problem solving
   forming, testing and revising a hypothesis
   confirming evidence
   prediction

   activities: choose variables
   gather data and set up data tables

3. Theory construction: experiments with natural language phenomena. In this component, students use their own language as the subject matter of scientific investigation. They formulate rules to account for certain regularities in their language. For example, to explain how regular English plurals are formed, students gather data, construct and test a hypothesis to account for the data, look for counterexamples, and revise the hypothesis until their

rule makes accurate predictions for speech. In this way, students are directly engaged in the process by which scientific knowledge is built up. The following content is covered:

terms: hypothesis, counterexample, theory, data

concepts: forming, testing, falsifying and revising a hypothesis
- the logic of problem-solving
- search for explanatory principles, theory-building

activities: pose an experimental question
- data collection
- communicate results to others

Overview of Piloting

The piloting of these materials was carried out during the months January through May 1985 with three groups of students from the Cambridge, Newton and Watertown school systems. Pilot sessions were held after school, once a week, for seven to eight weeks with STAMPS researchers and the students' science teachers.

The students in Groups 1 and 2 were drawn from grades 9 and 10, and in Group 3, from grade 7. Groups 1 and 3 had seven students each; Group 2 had five students. The software components were piloted with six subgroups of students: two subgroups of two to four students from each of the Groups 1 through 3. We divided each group into subgroups so as to keep the numbers small for working on the computer. Three to four sessions were spent on the software components: one or two on The King's Rule, two on The Scientific Method. Four linguistics sessions were held with the entire group in each school. Group 1 completed the linguistics sessions first, followed by the software sessions. This order was reversed for Groups 2 and 3. (See, Appendix A for details of the piloting schedule.)

The piloting was designed to collect information on the basic comprehensibility and potential effectiveness of the materials and lessons for the purpose of revising the materials, and for surveying the nature and range of children's ability to account for data as they formulate hypotheses. Thus, it was sufficient at this point in our work to collect and analyze data at the group rather than the individual level.

Plan for this Report

This document reports on this first round of piloting. The report also includes suggested revisions to the materials based on the pilot testing.

In what follows, we first describe the results of the pilot sessions for each of the components tested: The King's Rule first, The Scientific Method second, and then, the language problem sets. We then suggest an order in which the components of the curricular Unit should be taught. We discuss the competency levels shown by the students in the pilot sessions. And finally, we sketch in the work to be done in the third year of the Scientific Theory and Method Project.

PILOTING THE COMPONENTS

HYPOTHESIS FORMATION AND TESTING: THE KING'S RULE

Description of the Program

The STAMPS group chose The King's Rule as a piece of software to be used in the scientific theory and method Unit because we thought it could be useful as an introduction and exercise in hypothesis formation and testing, recognizing counterexamples, and hypothesis revision. Specifically, we felt the program would involve students in collecting data and, on the basis of an observed pattern, making hypotheses about how that data can be derived. In the face of conflicting data, students can revise their original and subsequent hypotheses until they feel they are correct.

The goal of the program is to advance through six levels of difficulty in hypothesis formation. The program takes the form of a game in which students move from room to room inside a king's castle by solving number problems, i.e., working out the rules that generate sets of number triads. Students progress from one room to the next, i.e., to a higher level of difficulty, by solving three consecutive problems in a given room. The objective is to reach the final room where the King resides, to solve the King's three challenging problems and thus, win three crowns. The program requires sixth grade math skills.

For each riddle, students are given three numbers related by a secret rule. Students must hypothesize what the rule is, and then test their hypothesis by entering three numbers of their own to see if the numbers fit the rule. Students can do this as many times as they like. When they are satisfied that they have discovered the secret rule, they can test their conclusion with a five-item quiz. The quiz consists of five triads which students must identify either as fitting the rule or not fitting the rule. If students respond correctly to all of the items, they pass the quiz and can move on to another problem. But if they incorrectly judge a triad they fail the quiz. They must try again at the same level. Passing three quizzes, i.e., solving three problems, at a particular level allows the students to progress to the next level of difficulty. To complete the program students must successfully complete three number problems at each of the six levels. (See, Appendix B for a detailed description of each of the levels of The King's Rule.)

For example, suppose the program gives the triad 18-6-12. The students first guess that the rule is <u>the first number minus the second equals the third</u> (rule name as given in The King's Rule). They have formed a hypothesis. To test it, they might ask whether 9-3-6 fits the rule. They are told "yes"; they have confirming evidence to support their hypothesis. They then proceed to test 21-7-14 and are told "no"; this triad is a counterexample. They must reformulate their hypothesis so that it generates 18-6-12 and 9-3-6, but not 21-7-14. The students recognize that the two triads that fit the secret rule share another common property: <u>all numbers are divisible by 3</u> (Level 3 rule). They seek further evidence by trying 21-30-3 and 90-60-54. Both of these triads are confirmed as fitting the rule. The students continue on to the quiz, and find that their hypothesis apparently matches the secret rule.

At the higher levels in the program some rules are quite opaque.

Transparent rules are those for which the number triads presented by the program clearly suggest the rules that govern them. For example "2-4-6" suggests the rule increasing jumps of 2 (a Level 1 rule). A less transparent rule governing this triad might be any three even numbers (Level 5); an opaque rule might be at least one number is divisible by 4 (Level 6). When the rules become less transparent, it is especially important to carefully test several different hypotheses so as to eliminate the incorrect ones. At the higher levels the rules are generally not the obvious ones and students must be on guard not to be misled.

Thus, it may happen that students have confirmation that they have discovered the secret rule when in fact they have not. Take the above example. When the program presents the triad 2-4-6, students might hypothesize that the rule is increasing jumps of 2. To test this, they could enter 5-7-9. If indeed this this triad fits the secret rule, they are told "yes"--even though the secret rule could be different from their hypothesis, e.g., any constant difference between the numbers (Level 5). At this point, unless the students deliberately test their hypothesis against this possibility, e.g., by testing sets of numbers that ascend or descend by some number other than 2, they will not get disconfirming evidence until the quiz.

Here is where the value of the program lies for our purposes. Students come to recognize that a given set of data may be consistent with more than one hypothesis, and they can try new examples to help decide among competing hypotheses. They learn to generate alternate hypotheses and to suggest triads that will eliminate incorrect ones. They become familiar with the terms "hypothesis" and "counterexample", and with the operations of testing hypotheses, finding confirming evidence or counterexamples, and revising hypotheses as necessary. They get experience in testing a new hypothesis not just against the latest data, but against all the data collected so far.

These operations and terminology provide a good basis for the work with real-world data and experimentation which follows in the Unit. Although the limited puzzle-solving nature of this activity and the artificiality of the domain might seem a detraction in terms of coming to understand the larger world of science, we feel that these very limitations provide a certain benefit for beginning students. They learn the concepts in a simple, contrived environment, and are ready to use this knowledge as they progress to studying the real world.

Results of the Piloting Sessions

Typically, when students started playing The King's Rule, beginning at Level 1, they would make a hypothesis after only one or two trials, and then go directly to the quiz for that rule. The rules were very easy -- simple addition and subtraction rules. They would usually get the quiz right because the rules were indeed simple and transparent. For example, the program on the first try in one subgroup gave them the triad 6-7-8. They hypothesized that the rule was ascend by 1 and tested this hypothesis with the triad 1-2-3. When the program responded "yes", they went directly to the quiz. In most cases, as in this one, they were correct. This was a common experience in all six of the subgroups at the first level.

But as the students moved on to Levels 2 and 3, they would find more often that their original hypothesis was not correct. The students learned to entertain different hypotheses and go through a number of trials. Thus each subgroup would get into the habit of generating several hypotheses and suggesting triads to decide among them. They were cautious about going on to the quiz, saying things like "Maybe we should try another one (triad)", "Let's try another one just to make sure", "I think we better try a couple more until we get the hang of it", and "You want to try a couple more just to make sure?"

For example, when the program presented the triad 3-9-27, one subgroup first guessed the rule multiply by 3, ascending, and tried 5-15-45. When the program responded "yes", they decided they had better chec ·hether the rule included descending order as well, and tried 18-6-2. A "no" response allowed them to reject the hypothesis multiply or divide by 3, and they proceeded to the quiz. In this case their testing proved adequate; the rule was multiply by 3, ascending.

The program helped the students recognized that it was necessary to think about whether the data were compatible with more than one hypothesis. They decided this was necessary because often the rules at these levels were less transparent, and they found they were being misled. Although they tested their hypothesis with a number of trials, each one confirming the hypothesis, they discovered this method of confirmation was not fool-proof. The correct rule might be one that generated all the data they had thought of, as well as triads they had not thought of. These other triads required a revision of the rule to cover both the old and the new data.

However, some students became impatient with the difficulty of the rules at Levels 5 and 6 and felt the program was unfair. "The rule could be almost anything!" was one reaction. We agreed that their frustration was justified, and have concluded that for our purposes it is enough to use the program through Level 4.

Of particular interest were the methodological errors some students made. For example, there was the 'consider the last example only' error, which arose when students met counterexamples in the quiz. The quiz presented a triad which their rule did not predict and so they rejected it. But the correct answer was "yes". Here was a counterexample, a triad which their rule did not predict but which needed to be included in the data. What they did, unfortunately and repeatedly, was redo their rule to fit only this example, not bothering to take into account all the preceding data.

A case in point was the subgroup that formulated a rule to cover 10-9-5: subtract 1 and then 4. When given 12-10-6 in the quiz they said "no". When informed that this was the answer was "yes", they decided to try the rule subtract 2 and then 4 (appropriate to the last example only). To the next quiz item 8-7-5 they said "no", and again were informed that they were wrong. They then tried the rule subtract 1 and then 2, still considering only the last example. These students were fully aware they were trying to derive the same rule, but it took time and experience for them to learn to revise a rule by taking into account all the data (rather than to just create an ad hoc rule to account for just the most recent piece of data). This was accomplished by encouraging the students to keep a complete record of the data so that it could be easily reviewed.

Another interesting error was the reaction of some students that the computer was making a mistake when it gave an answer that disagreed with their expectations. This was the 'only one rule works' problem, perceived as the computer going haywire. For example, one subgroup hypothesized that the rule for 4-12-36 was add 8 then add 24, and suggested the triad 3-11-35. When the program said "no", their reaction was "But the first you go up 8 and the second you go up 24", and they tried again. This time they tried 6-14-38. Again, the program said "no". They still stuck with their hypothesis and attributed the problem to the computer, asking "Why's it doing it wrong?" It was hard for them to realize that a triad or group of triads could be consistent with more than one rule. Since 4-12-36 is described by the rule add 8 then add 24, the students felt this rule must be correct and data to the contrary must indicate a computer error. Eventually these students caught on to the idea that a set of data could be consistent with more than one hypothesis, and stopped persisting with a rule in the face of counterexamples. They were then able to think about alternate hypotheses and what sorts of revisions to try.

Finally, we found that some students were unable to handle the arithmetic necessary to work through the program. Although the arithmetic needed is not above sixth grade level, one subgroup of 9th and 10th graders in particular could not handle the necessary division, and seemed unaware that division is the inverse of multiplication. They therefore could not figure out how to construct triads to test a rule they had formulated. Whereas the other errors described above are methodological errors central to what the program is trying to teach, this last one is extraneous.

Evaluation of the Program

Overall, we felt that The King's Rule was quite successful in introducing the students to the concepts and vocabulary associated with hypothesis formation and testing: hypothesis, trial, counterexample, hypothesis revision, prediction, confirming (supporting) evidence, disconfirming evidence. Students were afforded good practice in forming and testing hypotheses, and revising them as needed. We were able to introduce the terms "hypothesis" and "counterexample" as the students worked. These terms were understood in the context of the students' activities and in turn used by the students. Students learned to record data, and to consider all of the data when counterexamples were found and their hypothesis had to be revised. In addition, students learned to generate and test alternative hypotheses when attempting to determine rules.

Thus our first results are quite positive. We hope to secure permission to modify The King's Rule so as to exclude the opaque rules of Levels 5 and 6, and to include only Levels 1 through 4 (excluding a few opaque Level 4 rules). We fully expect this program to be part of the Unit that we have in preparation.

EXPERIMENTAL METHOD: THE SCIENTIFIC METHOD

Description of the Program

The STAMPS group considered The Scientific Method program valuable because it could provide an introduction to experimental method via the study of simulations of natural phenomena. We found it superior to other programs dealing with science methods in the material covered and in the way in which concepts are introduced, the practice it affords students, its pace and appeal.

The program introduces students to a "Scientific Problem Solving Guide". The guide is simply the five general steps one may follow in solving a problem scientifically:

1. Define problem
2. Collect information
3. Make hypothesis
4. Check hypothesis
5. Reach conclusion

The students work through a number of problem-solving games which highlight different features of this guide. Students also work through experimental exercises with the help of the guide. They learn to state a problem clearly and concisely, to identify the variables involved in a problem, to formulate a hypothesis by making an educated guess based on available data, to test the hypothesis by running a controlled experiment (with controlled variables and sufficient trials), to carefully collect all available information, and to draw conclusions based on the data at hand.

We adapted the program slightly before piloting it, to fit our requirements more closely. We replaced certain terms in the program to keep the terminology uniform across the curricular Unit. For example, where the original program used the terms "variable factor" and "control factors" for the parameters that an experimenter varies or keeps constant in an experiment, we inserted instead "independent variable" and "controls", terms which we are using elsewhere.

Moreover, before piloting we identified points in the program where it was necessary to stop to engage in group activity or discussion to enhance the program. Teacher intervention is necessary in order to have students explain and reflect on their thinking. We call these stopping points "exits", and intend to make these exits an integral, planned part of the program's use.

For example, an exit was required for students to better understand the set-up of a particular experiment, the Metal Containers activity. The major difficulty in this exercise for the students (as well as for the STAMPS group) was in identifying the number of controls. Here is the information given on the screen:

> The experiment is to determine what variable will cause the greatest amount of liquid to evaporate. Three containers are used. Each has a diameter of 21 centimeters. Each container is filled with 100 milliliters of boiling water. The first container is steel, the second is brass, and the third is aluminum. The three containers are allowed to sit for 20 minutes.

The question is, "How many controls are listed in this experiment?" Diameter, volume and time were fairly easy to identify as controls. But type of liquid (water) and temperature (boiling) must also be identified as controls. This question was never answered correctly in our pilot sessions. We resolved the problem by stopping and helping the students identify the intended controls. (We find the problem to be poorly worded, and may decide to change the wording, rather than leaving it as is with the need for an exit.)

Below we give examples of the exercises in the program and a brief description of their instructional value. We mention these five as being representative of the many activities on the disk.

Activity:	Instructional Objective:
Marble Problem	Think through a problem before coming to a conclusion.
Bermuda Grass Experiment	Introduction and definition of terms; give example of how to test a hypothesis by running an experiment; explanation of need to vary only one experimental factor at a time.
Metal Containers	Identify independent variable and controls.
Orange Experiment	Manipulate and test variables; discern and use relevant information; make a hypothesis based on data; decide whether results of experiment support or disconfirm the hypothesis.
Cricket Experiment	Take relevant notes on a phenomenon; formulate a hypothesis; simulate an experiment by choosing independent variable and controls and running a number of trials; record data; reach a conclusion based upon the data.

Results of the Piloting Sessions

We will discuss in detail the piloting of three of the above exercises: the Marble Problem and Bermuda Grass activity (both done in the first session), and the Cricket Experiment (the final exercise on the disk, done in the second session). Recall that piloting The Scientific Method extended over two sessions for all six subgroups.

(1) Marble Problem

The Marble exercise is a logic puzzle. It presents three closed boxes each containing a pair of marbles. The three pairs are: one black and one white marble(BW), two black marbles (BB), and two white marbles (WW). Each box is labelled (BW, BB, WW), but the labels are on the wrong boxes, i.e., the box labelled WW does not contain the two white marbles, and so on. The student, by looking at only one marble from one box of his/her choice, tries to correctly switch the labels so that they will match up with the marbles in the box. Using logic, it is possible to determine the arrangement of all the marbles by looking at only one marble. Which box should the student choose?

Students often quickly chose a box without thinking the problem through to examine the consequences of their choice. This exercise demonstrates what it means to think through a problem and trace the consequences of a decision.

We chose this point in the exercise for an "exit". When students chose a box we asked them to stop before pressing the key of their choice, and figure out what would happen under all possible circumstances. We wanted the students to try to figure out the system by thinking through the problem. We asked them questions such as:

"Suppose you get a white marble from the box you have chosen? Would that give you enough information to know for sure what's in each box? Suppose you get a black marble? Does that give you enough information?" The point of the exit was to distinguish logical problem-solving from a trial and error approach. We wanted students to understand that some problems are amenable to thought solutions and that problems ought to be examined for this possibility. Indeed, the Marble Problem is solvable by seeing only one marble as long as the right box is chosen.

It was interesting to observe the differences among students in their ability to think through this problem before actually pressing a key. Many students had trouble with this exercise in logic. But it was not only the logic which eluded them. Some found it a new experience to work through alternatives abstractly; this is the where we see the value of the exercise. This little logic puzzle gave students a chance to trace different pathways by thinking only, which required patience, thoroughness and accuracy. While not all of them could do this readily, this activity provided a starting point.

The transcript of a typical exit discussion provoked by this problem is included in Appendix C.

(2) Bermuda Grass Experiment

The Bermuda Grass problem presents a model of specific concepts involved in scientific methodology. It introduces and defines the terms "independent variable" and "controls", and shows how to run an experiment to test a hypothesis. It emphasizes the need to vary one parameter in an experiment (the independent variable), while keeping all other parameters (the controls) constant.

The Bermuda Grass problem involves a simulated experiment designed by a student testing whether an attitude of love or hate towards Bermuda grass will affect its growth. A picture of the grass shows that indeed, as the student hypothesized, the grass which he hated did not grow nearly as much as the grass which he loved. But when his teacher asks whether he watered the two equally his reply is, "Why should I water the grass I hate?"

This problem is a good (and humorous) example of the fault of not controlling variables. We had an exit here to discuss this problem. In the exit, every subgroup was able to recognize how the experiment was set up incorrectly and to understand the need to examine one variable at a time to obtain credible results. This experiment and the exit helped to instill the concept of needing only <u>one</u> independent variable and the problems that arise if you have more than one.

The students did this problem in the first session, along with another similar problem, the Metal Containers activity. Because of our scheduling, there was usually a delay of a week or more before the second session with The Scientific Method took place. The instructor ordinarily began the second session by asking if the students could recall what the terms "independent variable" and "controls" meant. In most cases, given the delay, the students could not recall the meaning of these terms. But when they were reminded of the Bermuda Grass and Metal Containers problems, they were able to recall the concepts involved.

For one subgroup, three weeks passed between the first and second sessions with The Scientific Method. A transcript of the discussion which opened their second session is given in Appendix D. That particular discussion is a good example of what we found more generally in using The Scientific Method program, and shows what we consider to be one of the program's greatest values. The activities give students specific and realistic models from which to understand and recall the concepts and vocabulary we are trying to introduce. Although they may not remember the terms after some time goes by, they remember the examples and concepts and can be reminded of the terms as needed. Thus, we expect teachers will be able to make use of these exercises and models in referring to the concepts and vocabulary in later science lessons after this particular Unit is completed.

(3) Cricket Experiment

The Cricket Experiment is the final exercise of the program and thus utilizes all of the concepts and vocabulary introduced in the program. It is presented in such a way that the students may themselves manipulate variables in a simulated experiment about what affects the chirp rate of crickets.

The menu for this activity looks like this:

Menu

1. Helps and Hints
2. Definition of Problem
3. Cricket Information
4. Make Hypothesis (Pick Independent Variable)
5. View Experimental Data
6. Perform Experiment
7. Reach Final Conclusion
8. Exit Program

The menu represents a procedure normally followed in performing an experiment: first, define the problem; second, gather information that might help in the experiment; third, make a hypothesis, etc. Students are allowed to choose any item on the menu; however, most subgroups followed the order of the menu. One subgroup went directly to the exam which they did not pass, and then returned to follow the order of the menu.

Here is a description of each of the relevant menu areas:

1. Definition of the Problem

This area defines and explains the problem which is to be solved. The problem is this: On three separate days while taking a hike around a pond a student has come across the same noisy cricket, chirping at a different rate on each of these days. Those dates and the corresponding chirp rates are as follows:

March 21	6:00 AM	Slow
June 21	2:00 PM	Fast
September 21	8:00 PM	Medium

This seems interesting and so the student decides to capture the cricket and run a series of experiments to try to figure out what is determining the chirp rate of the cricket.

The program at this point also asks: "Which of the following would be a good statement of our problem?"

A. Crickets seem to chirp at different rates.

B. Is there a single variable that affects the chirping rate of a cricket?

C. I wonder why and how a cricket makes his chirping noise.

Every subgroup but one picked B, the preferred answer in the context of the program; that subgroup chose A and on explanation of the answer quickly understood their error.

2. Cricket Information

This area gives information from books and articles about crickets which might be useful in making a hypothesis.

All the subgroups considered only two hypotheses about the chirp rate, basing them upon the information provided:

(i) From the information that only male crickets chirp and the fact that chirping is the calling song of the male, students thought that the rate might be related to the number of crickets in the area. The more crickets in the area, the more chances for the male to attract more females; thus, the faster chirp rate.

(ii) From the information that crickets chirp by rubbing one wing against the other, the students judged the dates and chirping rate to mean that the hotter the temperature the faster the crickets would rub their wings, in order to cool themselves off. This judgment would lead them to the correct inference, that chirp rate is affected by temperature. (The cricket's intentions remain open to question.) In this case, the students noticed the relevance of the date and time of day to higher or lower temperatures.

3. Make Hypotheses (Pick Independent Variable)

Having defined the problem as "What is the one variable that affects the chirping rate of the cricket?", students now select, from the list below, the independent variable which they think causes the cricket to chirp at different rates.

1. Humidity

2. Atmospheric pressure

3. Air temperature

4. The number of crickets nearby

5. Wind speed

This multiple-choice format deprives students of the chance to freely formulate hypotheses on their own. Thus, before looking at the given choices, there was an exit where students discussed the many hypotheses that could be made.

This format also restricts the students' understanding of what it means to formulate a hypothesis. Forming a hypothesis is not simply identifying a variable. The program fails to introduce the notion of relationship (e.g., inverse or direct) between an independent and dependent variable. After one of the above variables is chosen, there should be another exit where students discuss their predictions about the relationship of the independent variable to chirp rate.

Half of the subgroups directly acknowledged that the independent variable was most likely to be temperature, noting the relationship of the dates and times given to differences in seasons and temperatures throughout a day.

Many of the students explained their hypothesis in just this fashion, adding that the hotter it is the faster the crickets probably wanted to flap their wings to cool themselves.

4. View Experimental Data

This section allows students to view the experimental data according to the independent variable they choose to test.

Control settings for all variables are as follows:

Humidity	50%
Atmospheric pressure	29.2 inches (760 milliliters)
Air temperature	25 degrees Celsius (77 Fahrenheit)
Wind speed	5 m.p.h. (2 meters/second)
Number of crickets nearby	5 crickets

The test values can be set to any number within these boundaries:

Humidity	10 - 80%
Atmospheric pressure	10 - 90 inches
Air temperature	0 - 80 degrees Celsius
Wind speed	1 - 20 m.p.h.
Number of crickets nearby	1 - 20 crickets

5. Perform Experiment

This section allows students to run as many tests as they like, varying the setting of the independent variable and keeping all controls constant. For each setting of the independent variable that students choose, with all other control values being the same, the program provides the number of chirps/minute that the cricket chirps.

As mentioned above, half of the subgroups quickly acknowledged temperature as being the most probable cause for variations in chirp rate, given the variations in dates and times. These subgroups obtained immediate positive results, confirming their hypothesis in this experiment section.

The method all the subgroups used to test their hypotheses was to test the lowest and highest value of their independent variable first. Thus the temperature subgroups confirmed their hypothesis on this first round, and the other subgroups, using a different variable, disconfirmed their hypotheses. About half the time the subgroups also tested an experimental value in between the two extremes.

We recommend a graphing exit at this point in the program; we plan to have this exit in our next round of piloting. Students should stop and graph their results. It is appropriate here to introduce the notion of prediction. Students should draw a graph, get values from the program and enter points on the graph. Drawing a curve will allow them to predict a value for chirp rate for any temperature of their choice. They should then enter the temperature into the program and request a chirp rate. If the chirp rate matches their prediction it is confirming evidence that their hypothesis is correct and their procedures adequate.

The only otner variable tested was the number of crickets. This, of course, was always disconfirmed in the experiment.

Once students tested air temperature and obtained positive results, they would always try the quiz. They would test other variables (after obtaining the air temperature results) only after incorrectly answering the quiz. Wind speed and humidity were hardly ever tested except when students failed the quiz.

6. Reach Final Conclusion

Here the program tests the students with a five-item quiz based on their hypothesis of what affects the chirp rate (i.e., the independent variable they chose to test).

If any of the questions are answered incorrectly the program informs the students that they have not reached the correct conclusion and directs them to return to the menu and work through some further experimentation.

If the students answer all the questions correctly then the program responds with: "Congratulations!!! You have come to the right conclusion. It is definitely the temperature which affects the chirping rate of a cricket. As a matter of fact, in 1897, professor of physics A. E. Dolbear wrote an article called "The Cricket as a Thermometer"." The program then goes on to explain Dolbear's findings and presents the students with the Golden Cricket Award.

Five of the six subgroups (with three of the five testing air temperature first) went immediately to the quiz after obtaining the results of the air temperature experiment. All of these subgroups answered the following question incorrectly:

Question #2: If you double the temperature [independent variable tested], what will happen to the cricket's chirp rate?

A. It will double.

B. It will be halved.

C. It will stay the same.

D. None of these.

They chose A, for which they happened to have no evidence, and which is incorrect. In each case, the instructor had to point out to the students that they had not shown that the rate doubled when the temperature doubled. This was always confirmed by analyzing the experimental results of the extremes of temperature.

The fact that the students answered this question incorrectly indicates that the multiple-choice format drives the students to avoid the none-of-these answer. But in fact, there is a far more important criticism of this question; for it is based on an artifactual property of the Fahrenheit temperature scale. For example, when the temperature rises from 35 degrees F. to 70 degrees F., there is no real sense in which the temperature has doubled. What the question really addresses itself to is whether fixed increments of temperature change result in fixed increments in the chirp rate. Putting the question in this way points to the need to graph the results during the "Perform Experiment" section of the program. As discussed earlier, we plan to employ graphing in the next round of piloting. And we plan to modify Question #2 of the quiz accordingly.

The quiz also asks questions about whether changes in humidity, number of crickets nearby or wind speed affects the chirping rate of crickets. Each of the subgroups answered these questions correctly, eliminating these other variables--even if these variables had not been tested. The students thought that if air temperature affects the chirping rate then no other variable could affect it as well. The program sets up the students to take this for granted since it poses the central question of the experiment as "What one variable affects the chirping rate of crickets?". Thus, if they have already confirmed that air temperature does affect the chirping rate, then they should correctly conclude that no other variable can.

Evaluation of the Program

In sum, the program engaged students in a range of activities leading up to a simulated experiment about the chirp rate of crickets. We found each area of the program to successfully engage students in learning about the methods of science. In fact, most of the students thoroughly enjoyed the program, many of them saying that the program was very useful in understanding scientific method.

The various activities described above (e.g., Marble Problem, 3-Digit Game, Bermuda Grass Experiment, etc.) provided models for understanding of the methods of science. Whenever the instructor wished to discuss a concept or expand on a concept associated with the scientific method, he would usually refer to one of the activities to help the students recall what they had learned. The Metal Containers problem helped the students recall the definitions for "independent variable" and "controls". And the Bermuda Grass Experiment helped them recognize the importance of testing only one variable at a time.

The program offers students the flexibility to engage in any activity or area of the program in any order. They are able to exit at any time and move to a different section. The program also gives students the opportunity to choose among hypotheses, selecting independent variable values, and carry out as much testing as they like in an idealized (i.e., simulated) setting.

There are, however, certain limitations in hypothesis formation potential. The program equates hypotheses with independent variables. For example, in the Cricket Experiment only five hypotheses, defined as variables, are offered in a multiple choice format. The program offers no opportunity for students to speculate freely about other possible variables (e.g., wind direction) or to specify the relationship between the independent and dependent variable (i.e., does chirp rate increase or decrease with a rise in temperature?). We feel this problem can be dealt with in the Cricket Experiment by introducing exits, encouraging students to thoughtfully formulate their hypotheses.

Another limitation concerns the choice of independent variables in an experiment. The program does not permit students to work with more than one independent variable at a time. In the Bermuda Grass Experiment the program explains why having more than one independent variable is unacceptable procedure, but it does this in a stipulatory fashion and not in a discovery fashion. There are other programs which allow students to choose as many independent variables as desired in performing an experiment, usually leaving students to find out for themselves that having two independent variables is inadequate. We are presently considering what means to use to permit discovery of this point.

A final point to mention is that much of the vocabulary in the program was not easily definable by students later in the piloting. They could not define, for example, "trials", "counterexample", "independent variable", and "controls". However, we are confident that they understood the concepts because discussion of the model cases showed that they could operate with the notions. In discussion of the Bermuda Grass or Metal Containers problems, students were able to easily identify the need for having only "one thing that changed and all the others the same". When we reminded them of the terms "independent variable" and "controls", they were able to successfully use the vocabulary. What apparently will be necessary in the future for both understanding and retention of the vocabulary is structured use of the terms in exits, exercises and reviews.

THEORY CONSTRUCTION: THE LINGUISTICS COMPONENT

Description of the Linguistics Component

In the linguistics component of the Unit students use their own language as subject matter for scientific investigation. They derive certain rules of English to explain aspects of their own speech. This part of the Unit gives them an opportunity to work with real-world data that they collect themselves, and to figure out principles which account for the facts they have observed. The accessibility of language data and the ease of working with such familiar material combine to make language an attractive candidate for this beginning study of scientific method.

Students actually construct scientific theories -- small ones, to be sure, but theories nevertheless. They use data from their own language to derive linguistic principles for particular structures. Students engage directly in the actual methods used in conducting scientific research, namely recognition of a phenomenon to be explained, data collection, data classification, and hypothesis formation. They test their hypotheses, look for counterexamples, and revise the hypotheses until they account for the data and can predict accurately. This first-hand experience in actually constructing a scientific theory helps students understand the abstract nature of a theory, and the relation between theoretical principles and the facts they are constructed to explain.

The idea of "doing linguistics" as an introduction to scientific method and theory construction has been suggested before (see excerpt from Hale, 1975 in Appendix E). However, this approach has not to our knowledge been tried with secondary school students. Our piloting has been very much an exploration of what structures to use, and how to present them. Our questions have been: Can the students generate the data and make the linguistic judgments readily? Can they work out the principles?

Results of the Piloting Sessions

We worked with the three groups in succession, completing the work with each group before starting with the next. We were thus able to analyze the results from each group, decide which constructions we wished to repeat or discard, and improve our methods of presenting the successful structures.

In all we tried seven constructions; three of these turned out to be quite successful: plural formation, want + to contraction (wanna), and x + is contraction (John's late).

We found that students understood the problems, had no trouble generating the data, and were able to make the necessary linguistic judgments easily. They formulated rules, looked for counterexamples, and revised their rules as needed. With help and discussion, they worked out principles that explained the data. We are encouraged by these results, and plan to keep the linguistics component in our final curricular Unit. We are working on improvements for the next round of piloting.

Below we discuss in detail the sessions with Groups 1 and 3. The discussion reveals the different approaches and responses of both the instructors and the students towards the problems. We begin with an account of the third and final group of students; five 7th grade students participated in each of these sessions.

Note that in the following discussion, we indicate sounds by enclosing letters in slash marks. An asterisk (*) indicates a construction which is not well-formed.

Group 3, Session #1: Five students participated in this session: J, R, G, B and K; all had previously participated in the software sessions. The instructor worked on the blackboard. Two observers were present.

The instructor began with a general introduction to linguistic phenomena, one of which was plurals. The session then focused on formulating a hypothesis to account for regular plurals.

--The instructor asked students for their judgment of: I saw five *dog/s/. Of course, they didn't accept this. He asked them to figure out a way to explain to a Martian how plural forms are made.

--B gave a description based on spelling:

(a) Add s, es, ies.

--The instructor reminded the students to pay attention to how plurals sound because the Martian would want to know how to say plurals.

--The students generated data with the full range of plural endings (i.e, /iz/, /s/ and /z/). K made the following hypothesis which accounted for only some of the data:

(b) Add an s.

The instructor reminded the students to focus on the sounds of the plurals not their spelling. He introduced slash marks as a convention indicating sounds. He wrote the rule as:

(b') Add a /s/.

--This rule (b′) was tested and found inadequate. K suggested:

(c) Add a z sound.

This was added to (b′) on the board:

(c′) Add a /s/.

Add a /z/.

--The instructor pointed out a counterexample to (c′) in the data: house --> *house/z/. B refined the hypothesis to:

(c") Add a /s/ or /z/ depending on how the word ends.

This did not account for the counterexample.

--B, J and R reformulated the rule to account for the third plural ending (/iz):

(d) Add a /s/ or /z/ or /iz/ depending on how the word ends.

--The instructor asked them what the rule meant, i.e., what did "depending on how the word ends" mean? B began to identify the word endings requiring /iz/. He noted that /ch/ sounded "sorta like an s"; the instructor agreed that it was "s-like". The students generated more data with the /iz/ plural and identified other "s-like" sounds.

--One part of (d) was then made more specific:

(e) Add /iz/ if the ending of the singular word ends with an "s-like" sound.

--The students turned to specifying when /s/ or /z/ was added. They generated more data with these plural endings. They could not generalize across the data; B said, "There's no real rule."

--The instructor gave the students pairs of words which contrasted in the voicing of either their initial or final sounds, e.g., /p/uck - /b/uck; do/ck/ - do/g/. He asked the students how the pairs of sounds differed. B distinguished the voiceless sounds as "sharp, less vibrant". The instructor equated the descriptions "sharp, less vibrant"/"dull, more vibrant" with the conventional linguistic terms "voiceless/voiced".

--B revised part of hypothesis (d) and added it to (e):

(f) Add /iz/ if the ending of the singular word ends with an "s-like" sound.

Add /z/ if word ends with voiced sound.

Add /s/ if word ends with voiceless sound.

--Next, the instructor focused on the difference between /s/ and /z/. B identified /s/ as voiceless and /z/ as voiced. The rule (f) was refined as follows:

(f′) Add /iz/ if the ending of the singular word ends with an "s-like" sound.

Add voiced to voiced.

Add voiceless to voiceless.

--The hypothesis was tested. It was found to account for all prior cases and to make positive predictions.

--The instructor suggested the plural rule (f′) could be extended to possessive forms (Mike's book, John's book, George's book) and to contractions of is (Mike's late, John's late, George's late). The students generated data to test these suggestions. B agreed, "It's gonna be the same rule as before."

General comments:

This session went very well. The introductory remarks succeeded in capturing the students' interest in language phenomena; plural formation was not presented as an isolated phenomenon but as just one of several linguistic mysteries to be explained.

The students' science teacher, who was present at the session, noted with some surprise the active participation of B and R; neither boy does well in school. In fact, all the students took part in the discussion and seemed to follow the group's thinking.

At the end of the session, the instructor suggested two new constructions, possessives and contractions, with which students could test their plural hypothesis. Suggesting these constructions worked better than having the students themselves search for constructions which resembled plural forms, as we had tried earlier in other schools. Compared to the approach used in the other two groups, extending the plural hypothesis in this way seemed more useful.

Terms used: data/data base
hypothesis
counterexample
linguistic terminology: judgments, voicing,
possessive nouns

Discussion of: hypothesis formation
prediction
hypothesis testing
revising a hypothesis
linguistic symbol convention: / /

Group 3, Session #2: Five students participated in this session: J, R, G (from Session #1), C and M (new). The instructor gave each student a data-sheet. (See, next page for the Data Sheet.) He worked on the blackboard and the students recorded their work on the data-sheet. Two observers were present.

The instructor began the session with some remarks about linguistics and the nature of linguistic data. The session focused on formulating a hypothesis to account for want + to contraction into wanna.

--The instructor described linguistics and likened it the study of the sun where you can only study your observations of the sun's surface. He distinguished two kinds of data used in the scientific study of language: judgments and interpretations. He gave an example of what he meant by judgments, and introduced the asterisk (*) as the symbol used to mark ill-formed constructions. He then worked through section (A) on the data sheet to illustrate how interpretations (of how words work together) constitute data.

--The instructor asked the students whether want and to could be contracted to form wanna in the questions in section (B). The students decided contraction was possible in 1-4 (although C differed on 2) and not possible in 5-6. Questions 1 and 5 illustrate the contrast between the questions where contraction was possible and those where it was not:

1. Who do you want to speak to?

   wanna - ok

5. Who do you want to speak?

   *wanna - not ok

DATA SHEET

(A) Preliminaries

1. Alex is too stubborn to talk.
2. Alex is too stubborn to talk to.
3. Alex is too stubborn for us to talk to.
4. Alex is too stubborn for us to talk to him.
5. Alex is too stubborn for us to talk to Emma.
6. Alex is too stubborn to talk to Emma.

(B) Problem I: <u>want + to</u> contraction [=<u>wanna</u>]

1. Who do you want to speak to?
2. What do you want to win?
3. Where do you want to go?
4. When do you want to speak?
5. Who do you want to speak?
6. Who do you want to win?

(C) Problem II: <u>x + is</u> contraction [=<u>x's</u>], where x= anything at all.

1. Peter is short and Emma is tall.
2. Emma is tall and so is Peter.
3. Emma is tall and Peter is too.

4. I wonder when Emma is at home.
5. I wonder what Emma is looking at.
6. I wonder who he is sometimes.
7. I wonder where Emma is today.

8. Peter was happier yesterday than Emma is today.
9. Emma thinks; therefore she is.

--The instructor asked the students to state the problem of <u>want + to</u> contraction. Jennifer responded, "When do we use wanna?"

--The instructor introduced the term "wh-words" (called "question words" by the students). To focus the discussion, he asked, "Where do you interpret the wh-words?" He suggested the students try the "surprise question" tactic. A surprise question is one that indicates surprise on the part of the speaker, not just a neutral request for information. For example,

normal question: Who do you want to speak to?

surprise question: You want to speak to <u>who</u>?

In 1-4 the wh-word was interpreted at the end of the question:

1. You want to speak to <u>who</u>?
2. You want to win <u>what</u>?
3. You want to go <u>where</u>?
4. You want to speak <u>when</u>?

In 5-6 the wh-word was interpreted between <u>want</u> and <u>to</u>:

5. You want <u>who</u> to speak?
6. You want <u>who</u> to win?

--Given these observations of the point of interpretation, the instructor asked if a hypothesis could be made about when <u>wanna</u> was allowed. J offered the following:

(a) When talking about something you're gonna do in the future.

This hypothesis was immediately rejected since it didn't account for the data.

--There was some confusion about the word "interpretation", because the instructor was using it in more than one way (e.g.,to describe the mental location of a wh-word; to describe how words work together).

--The instructor then asked the students when contraction was not allowed. To focus the discussion, he asked them to compare the following:

i. I want Bill to go

ii. Who do you want to win?

iii. Where do you want to go?

They saw that want + to contraction was not allowed in the first two.

--The instructor restated the question, "Can you make a hypothesis about where contraction is prevented?" J hypothesized:

(b) When you got something between want and to you can't contract.

--The instructor specified "something" based on the observations and added to (b):

(b') When you got something between want and to you can't contract: where "something" is a word or a position of interpretation.

--Next, the instructor turned to C's judgment of 2 (which appeared to be a counterexample to the hypothesis (b')). The students determined it was not a counterexample since C was interpreting the wh-word between want and to, as if the surprise question was "You want what to win?" This raised the issue of structural ambiguity (which had been defined earlier). The instructor suggested a way in which 6 might also be ambiguous, which the students found amusing.

The discussion of 2 and 6 as possible counterexamples provided an interesting way to test the hypothesis.

--The instructor made some remarks about the process of science and about the hypothesis that had been reached. He pointed out that the <u>want + to</u> contraction showed there are things in language which act like words but which can not be heard in speech.

--The instructor asked the students for their judgments of <u>x + is</u> contraction in sentences in section (C) on the data-sheet. He said the discussion would be continued in the next session.

General comments:

This session went very well for the following reasons. The discussion of Preliminaries in section (A) of the data-sheet provided a good introduction to the problem of <u>want + to</u> contraction. Asking the students for their judgments--and pursuing them (e.g., C's minority opinion)--engaged them in the contraction problem. Also, the data-sheet provided a structured way for the students to work through the data and record the group's observations.

Terms used: data

hypothesis

linguistic terms: judgments, interpretations, point of interpretation, ambiguity

Concepts discussed: hypothesis formation

posing an experimental question

counterexample

hypothesis testing

explanation

linguistic symbol convention: *

Group 3, Session #3: Five students participated in this session: J, R, B, C and M. (J and R had participated in the two previous linguistics sessions; B, C and M in only one.) The instructor gave out clean copies of the data-sheet used in Session #2 (see above). He used the blackboard and the students recorded their work on their hand-outs. Four observers were present.

The instructor briefly reviewed the previous session's work. The session focused on formulating a hypothesis to account for x + is contraction.

--The instructor asked the students whether is could contract in sentences 1-9 in section (C) of the data sheet. The sentences and the students' judgments are shown below:

1. Peter is short and Emma is tall.

   Emma's - ok

2. Emma is tall and so is Peter.

   so's - ok

3. Emma is tall and Peter is too.

   *Peter's - not ok

4. I wonder when Emma is at home.

   Emma's - ok

5. I wonder what Emma is looking at.

   Emma's - ok

6. I wonder who he is sometimes.

   *he's - not ok

7. I wonder where Emma is today.

*Emma's - not ok

8. Peter was happier yesterday than Emma is today.

*Emma's - not ok

9. Emma thinks; therefore she is.

*she's - not ok

(The words targeted for contraction were not highlighted on the data-sheet; the data-sheet needs to be modified.)

--J suggested the following hypothesis:

(a) You only contract is after a noun.

She immediately rejected this hypothesis because of sentence #2.

--The instructor asked the students whether the word tall could be added to sentences 1-3. They noted the following:

(b) If tall is there we can contract; if tall is not there we cannot.

--The instructor turned to sentences 4-7 and asked the students to interpret the position of the wh-words. They noted that the wh-word was interpreted at the end of sentences 4-5, between is and sometimes in 6, and between is and today in 7.

4. Emma is at home when? can contract

5. Emma is looking at what? can contract

6. He is who sometimes? cannot contract

7. Emma is where today? cannot contract

--The instructor asked, "Do these observations (of the judgments and interpretations) tell us something?" He asked if anyone had a hypothesis. J suggested:

(c) You can only contract if it's at the end of sentence.

The instructor wrote the following on the board:

(c') Can only contract is if the wh-word is interpreted at the end of the sentence.

--The instructor asked if this hypothesis accounted for the data. The students thought it did. He asked them how they would test the hypothesis. The students discussed what would be a valid counterexample to falsify the hypothesis. Initially, B suggested that a counterexample would be a sentence where the wh-word was interpreted at the end. Later, he decided an appropriate counterexample would be a sentence where the wh-word was not at the end of the sentence but where x + is contraction was allowed. Rule (c') was then tested by lengthening a version of sentence 5:

I wonder what Emma is giving to Peter today.

Emma's - ok

J said this sentence did not fit the hypothesis.

--There was a long discussion of 4-7 which focused on the point of interpretation of the wh-word in each sentence. The students observed that the wh-word was interpreted at the end of sentences 4-5 and after is in sentences 6-7. Eventually, J made the following hypothesis:

(d) Can't contract if it's interpreted right after is.

The instructor wrote the following on the board:

(d') Can't contract if the wh-word is interpreted right after is.

--The hypothesis was tested against the data. Sentences 2, 3 and 8 revealed that the hypothesis was not general enough, because in these sentences the interpreted word is not a wh-word. B and J suggested:

(e) Can't contract if the word being interpreted/the interpreted word is interpreted right after is.

To be even more general, the instructor modified this and wrote the following on the board:

(e') Can't contract if something -- a word, wh-word -- is interpreted right after is.

--The hypothesis was tested against the data, and found to work for sentences 1-8.

--Sentence 9 was discussed. J said, "It's a strange sentence." The instructor talked about the role of stress in is contraction. None of the ideas raised in the discussion were incorporated into the hypothesis.

--The students evaluated the linguistic sessions. They understood that what they were doing was making hypotheses and testing them. They said the material was interesting and not too difficult. B said, "It's not really hard but it makes you think."

General comments:

The students were actively engaged in giving their judgments of sentences, thinking about the data and formulating hypotheses to account for x + is contraction. Although there were periods of silence, the students did not seem to be bored by the material; rather, they seemed to be thinking very hard.

The discussion of the exceptional nature of sentence 9 did not seem fruitful.

Terms used: data

observations

hypothesis

Concepts covered: counterexample

hypothesis testing

falsifying a hypothesis

revising a hypothesis

With Groups 1 and 2 the discussions took a somewhat different turn. With the set of 9th graders in Group 1, for example, discussion of the plural construction extended into two sessions, and the students were able to readily generalize aspects of the plural rule. Further, when given the opportunity to extend the principles to past tense formation, they were able to appreciate the similarities. It should be noted that the students in Group 1 are two to three years older than the students in Group 3 described above.

Group 1, Session #1: Five students participated in this session: M, J, A, L and N. The instructor had the students record the group's work on their own sheets of paper. There were three observers present.

The instructor introduced the problem of how regular English plurals are formed. The session then focused on formulating a hypothesis to account for such forms.

--The students were able to generate data and place them into two categories: forms ending in /z/ and /s/. The instructor introduced data which constituted a third category (i.e., forms ending in /iz/) which the students generally ignored throughout the session.

--Initially, the students formulated hypotheses which could account for only some of the data generated and which had limited predictive power:

(a) Words ending with -ook, -op and -at take s [e.g., <u>books</u>; <u>bats</u>; <u>tops</u>].

(b) Words ending in -ck, -lk, -ek take s.

(c) Words ending with -en and -y take z [e.g., <u>kitchens</u>, <u>pens</u>; <u>stories</u>].

Hypothesis (a) came about in part because the students generated words that rhymed (e.g., <u>books</u>,<u>looks</u>, <u>cooks</u>; <u>fats</u>, <u>cats</u>, <u>bats</u>). The instructor refined (b) to apply to words ending with k; this pushed students to consider other consonant endings.

--The instructor asked the students to identify the difference between /s/ and /z/. M felt the difference in her tongue and J said that with /z/ the tongue vibrates. The instructor then introduced the notion of voicing and voicing pairs, and went through a number of such pairs many of which M identified. The students then looked at the different categories of plurals to see if voicing (called "vocal cord vibration" by the students) mattered.

--J formulated a hypothesis (d) which covered two of the three categories of data (the two which the students had created):

(d) Anything that makes your voicebox vibrate (like b, d, g) gets a z sound. Anything that doesn't vibrate gets s.

--The instructor and the students went through their lists of plurals to test this hypothesis. It seemed to work; however, data of the /iz/ form were ignored (e.g., bushes, horses, lushes). The instructor introduced a counterexample: thrush/thrushes. The students then tried to revise the hypothesis to account for words that take /iz/ and came up with several ideas:

(e) Any word with sh [e.g., thrush] takes ez.

(f) Any word with se [e.g., horse] takes ez.

(g) Any word with the s sound (i.e., /s/) takes ez.

--The instructor introduced "hissing sounds". Eventually, J added a caveat to his earlier hypothesis (d):

(h) But, if it's close to an s or a hissing sound it takes ez.

--The instructor introduced the notion of ordering the different parts of the rule so that it would be more elegant in its operation. J reordered the steps in the rule so that it then looked like this:

(i) If it's close to an s or a hissing sound it takes ez.
If your voicebox vibrates it gets a z.
Everything else gets a s.

The science teacher pointed out that this was like writing an 'if...then' statement in BASIC (something which most of the students seemed to know about).

--The instructor discussed the application of this rule (i) to words ending in non-English sounds, e.g., German Bach. This was to illustrate that the rule operated independently of individuals' vocabulary.

--At one point in the session, the students initiated a discussion about what was meant by "doing science". The instructor explained that generating English plurals and formulating a hypothesis about how they are formed was what "doing science" was about: trying to explain a phenomenon given a set of data.

General comments:

Not surprisingly, at the start of the session students focused on the spelling of words and their plural endings rather than the sounds. This is a problem since the spelling of plural forms does not directly reflect pronunciation; for example: contrast girls and cats, where s represents both /z/ and /s/. Spelling conventions also affected the way in which sounds were represented in written form by the students. /s/ was written as s; /z/ was written as z; /iz/ was written as ez (and was often referred to as "the e-z ending"). This representation of /iz/ confused spelling and sound.

The students appeared highly motivated to pursue this problem. That they initiated a discussion about "doing science" indicated general curiosity about the project.

Terms and concepts introduced: hypothesis

counterexample

data vs. theory

linguistic terminology: voicing

Discussion of: hypothesis formation

hypothesis testing

revising a hypothesis

ordering in rule application

Group 1, Session #2: Five students participated in this session: M, J, A, R and D (M, J and A had participated in the previous session). In this session, the instructor worked on the blackboard. Two observers were present.

The instructor began the session by saying he wanted to do three things: (1) restate the hypothesis about plurals, (2) sharpen the hypothesis and make it better, and (3)extend the hypothesis if it had relevance to other things in English.

The instructor emphasized that the discussion had to do with sounds not with spelling. He introduced slash marks as a convention indicating sounds. (There was some discussion of the representation of /iz/ which had previously been called ez.)

--The students, led by J, restated the hypothesis:

(a) 1. If the word ends with a s-like or hissing sound, then add /iz/

2. If the word ends with a vocal cord vibration then add /z/

3. Elsewhere add /s/.

--A long discussion was initiated by the students about why the sessions were about linguistics and why the project was taking place. The instructor described linguistics and likened it to other sciences where we work with observable phenomena. He said the study of science, in general, involves both learning the facts of science and a way of thinking about things; we are emphasizing the latter.

--The instructor pushed the students to refine the hypothesis by looking for a more general operating principle. He focused the discussion by asking, "Why do you think things (i.e., plurals) should be this way?" The students came up with several ideas which led J to revise the hypothesis (a) to refer to the abstract feature of voicing rather than to specific sounds:

(b) So if the word ends with vocal cord vibration, keep vocal cord vibration.

If it doesn't, then don't.

--M noticed that the s-like sounds were not covered by this description. Since (b) was presented orally, it was not clear whether it replaced all or part of hypothesis (a). The instructor responded that an important part of the rule was ordering, implying that (b) had replaced the last two steps of (a).

--J revised the hypothesis further by collapsing the statements about voicing in (b) into one statement:

(c) Keep end sound going.

This is what is referred to in linguistic terminology as an "alpha rule". The instructor pointed out that this made the rule more economical. The rule as the instructor wrote it on the board looked like this:

(c') If the word ends with a s-like or hissing sound, add /iz/.

Choose the plural ending by the vibration of the last sound of the singular.

--The instructor asked the students if this rule (c') had relevance to other things in English: "Are there any other places to add s?" A suggested that the hypothesis be tested to see if it worked for verbs; following this suggestion, the instructor directed the students to extend the hypothesis to include 3rd person singular verb forms, e.g., runs. After a long discussion which revealed the students' confusion about the question, D suggested possessives. Students were not able to generate data.

--Because of the discussion of verb forms, there was an argument about renaming the plural hypothesis to either M's suggestion, "end with S-hypothesis" or J's "end-sound hypothesis". The argument focused on the generalizability of the stated rule (c'). The "end with S" label was agreed upon.

--The instructor gave the students several examples and asked them to think about contraction (i.e., of x + is, x + has) and regular past tense forms (e.g., wished, buzzed) for the next session.

General comments:

The students appeared genuinely curious to know why solving a linguistic problem was considered part of science rather than English (i.e., traditional grammar). In extending the plural hypothesis to other constructions in English the students brought traditional grammar terms--and their confusion about what they meant--into the discussion.

Terms used in the session: hypothesis
counterexample

Discussion of: explanation
revising a hypothesis
prediction
symbol convention
ordering in rule application
economy

Group 1, Session #3: Five students participated in this session: M, J, A, L and D. (M, J and A had participated in Sessions #1-2; L in Session #1; D in Session #2.) The instructor worked on the blackboard during this session. Two observers were present.

The session focused on three things: (1) a review of the S-ending hypothesis as it was formulated in Session #2, (2) extending the data, and (3) formulating a hypothesis to account for regular English past tense forms.

--The S-ending hypothesis was restated as follows:

1. If the word ends with a hissing sound, then add /iz/
2. Continue the voice-box vibration of the last sound.

The second part of the hypothesis was explained to L who was not at Session #2.

--The instructor quickly reviewed the extension of this hypothesis beyond plurals to include verbs and possessive forms.

--The students were able to further extend the data base covered by the S-ending hypothesis to include the contraction of _is_ (e.g., _judge's_: /iz/, _it's_: /s/, _he's_: /z/), _does_ and _has_. The students saw that these forms of contraction were distributed in the same way as the plurals.

--The instructor reminded the students about J's suggestion from Session #2 that the hypothesis might be more general (i.e., "the end-sound hypothesis"). He then introduced the problem of how the regular English past tense is formed. A and L quickly responded with a hypothesis:

(a) Add ed.

The instructor pointed out that _ed_ indicated the spelling rather than the sound of the past tense.

--The students then generated data, e.g., cried, stayed. M had another hypothesis which was specific to these data and had limited predictive power:

(b) It's really /d/. That's all it is--/d/.

--Almost immediately, however, M generated the word smacked, and noted the /t/. The students then continued to generate past tense forms and placed them into three categories: forms ending in /d/, /t/ and /id/.

--J initiated the "past tense hypothesis" to cover two of the three categories of data (i.e., forms ending in /t/ and /d/); with M's help, it was stated as:

(c) Keep the vibration because if that vibrates, take /d/ and if it doesn't take /t/.

--The students had some difficulty formulating a statement about the /id/ forms. They tended to focus on the type of vowel within the word (i.e., long vs. short). After the instructor focused their attention on the endings of the words, they were able to formulate a hypothesis with predictive power which could account for all of the data generated:

(d) If the word ends with t/d sound, then add /id/. Continue voicebox vibration.

--The students saw that the S-ending hypothesis and the past tense hypothesis had many similarities. The instructor talked about how the second part of both hypotheses (i.e., "Continue voicebox vibration") is very general in many languages.

--J visualized a diagram which abstracted this phenomenon (of voicing assimilation) which the two hypotheses accounted for:

(e) Put this hypothesis in a diagram. You could have a bigger circle being "continue voicebox vibration of the last sound", and then off of it you could have the S-ending, D-ending, etc., etc., etc.

--There was some discussion of whether to call the past tense hypothesis the D-ending or T-ending hypothesis. The instructor told the students it did not really matter.

General comments:

Considering there was a bomb scare (during the discussion of x + is contraction) and the building was evacuated for about 15 minutes, the session went well. The students continued to be interested and motivated by the linguistic material.

Terms used: hypothesis
data

Concepts understood: hypothesis formation
prediction
hypothesis testing
explanation
revising a hypothesis
generalizability

With Group 2, 9th and 10th graders in a third school, the sessions on plural formation were less successful. The students had difficulty forming workable hypotheses to account for the data, and displayed a lack of interest in the material. They were inattentive and bored, though polite. Their teacher noted that their lack of involvement was very similar to what she had observed in her classes with them. For students who tend not to be motivated by schoolwork in general, the linguistics examples may be inappropriate, since they assume a curiosity about how things work and a willingness to deal with abstract material.

Evaluation of the Language Problems

In the course of the pilot sessions, we have reduced the set of language problems that will be used in the Unit to the plural formation and contraction problems; extensions of the plural hypothesis and regular past tense formation may be used to supplement plural formation. The linguistic constructions that we tried and have rejected as being less successful are tag question formation and yes/no question formation. Let us now briefly justify this selection process.

1. Regular English plurals

This problem required students to generate data, to categorize the data and to describe the conditions for the distribution of the plural forms. In describing the distribution, students could move from a list approach to an analysis of the features of the sounds with which words end and their plural forms; this move involved generalizing across the data in order to abstract a feature of those data.

Asking the students to generate plurals immediately engaged the students' attention. This relatively simple task produced the full range of data in a short period of time. Identifying the underlying conditions for the distribution of plurals required students to abstract the features of "voicing" and "s-like sounds"; both of these features seemed quite accessible to most of the students.

It was hard to get students to focus on plural sounds rather than on spelling. An interesting approached was used with Group 3. The students were given the task of telling a Martian how to say plurals (i.e., how to sound like a human being). It soon became clear to the students that a spelling rule would not help the Martian.

It did not seem useful to spend a lot of time refining the hypothesis (as in the second session of Group 1). Such a process did not engage students in working through data.

2. Extensions of the plural hypothesis

The discussion of other constructions similar to plural formation was useful only as a way of testing the plural hypothesis.

When students were asked to provide constructions (as in Group 1), the discussion became tedious and frustrating. Students felt obliged to name the construction they were thinking of; this naming process was obviously a source of confusion for all of the students, since only some of them knew the appropriate traditional grammatical terms.

Providing examples of the related constructions (as was done with Group 3), focused the discussion more usefully on testing the plural hypothesis.

3. Regular past tense

The problem of accounting for regular past tense forms is quite similar in structure to plural formation.

The students in Group 1 were asked to generate and categorize the data, and to describe the conditions underlying the distribution of the past tense forms. Because of their experience in working through plurals, they immediately tried to abstract features from the data to specify these conditions.

4. Tag question formation

This problem required students to generate tags and to formulate a rule which would generate good tag questions. Describing the data generated involved specifying the verb form copied, the form of the subject copied, the location of the copied forms, the positive/negative variation, and the appearance of do forms. The descriptions were largely a listing of the forms and their locations. In describing the form of the subject and the do forms, the students had to consider more abstract aspects of the data.

The range of data considered was controlled by the use of a number of data-sheets. The process of formulating a rule was structured by the data-sheet(s) under consideration. Thus, each succeeding data-sheet provided counterexamples and required the rule to be either revised. After a while, this process did not seem useful; it was somewhat similar to the process of refining the plural hypothesis (done with Group 1), although the students did work through the data. It might have been more useful to have the whole range of data available from the start.

Confusion over grammatical terminology slowed the pace of the discussion. Although an account of tags need not refer to traditional grammatical terms, it inevitably does rely on such terminology.

5. Yes/no question formation

This problem is similar to the tag question construction. The students were asked to generate and account for a set of data. In formulating a rule for yes/no questions, the students described the type and location of the noun and verb forms, and the appearance of did and does. The account of do forms required consideration of an abstract aspect of the data.

As in the discussion of tag questions, the use of grammatical terms posed a problem. The clearest example of this was one student's inability to test the final hypothesis because she did not know what a verb was.

6. want + to contraction

This problem required students to generate a limited set of data (i.e., judgments and interpretations), to make observations of these data, and to abstract a condition blocking contraction.

Asking students for their judgments of contracted forms immediately engaged their attention. They clearly could make judgments of a very subtle and intriguing nature.

Abstracting the condition blocking contraction required the students to interpret the position of wh-words and to note that something not present in speech had the same effect as a word. This provided an interesting linguistic comparison to other phenomena studied by scientists (e.g., gravity).

With the Group 2, want + to contraction was considered after x + is contraction. Since want + to contraction is somewhat easier, this was not an appropriate order.

7. x + is contraction

This problem is similar to want + to contraction. Students were asked to generated a limited set of data (i.e., judgments and interpretations), to make observations of these data, and to abstract the condition blocking contraction.

As with want + to contraction, students could readily give very subtle judgments of contracted forms. They also had to identify the point of interpretation and its blocking effect. For the students from Group 3 who had already worked through want + to contraction, this task was immediately clear.

Although the data were considered in sets, the full range of data was available to the students throughout the session. This was helpful since it was clear what had been accounted for and what data needed to be incorporated into the description. The consideration of truly exceptional data was not useful.

In summary, we believe that we had the greatest success using the regular plural form problem and the contraction problems. The yes/no question problem and the somewhat overlapping tag question problem appear to be too dependent on tradional grammatical terminology that students do not know or do not want to learn. It may feel too much like ordinary school grammar slightly repackaged. On the other hand, the plural forms and the contractions involve phenomena that are never dealt with in school grammar apart from their spelling conventions. Once the students' attention is redirected toward the phonological and syntactic phenomena involved, the students proceed to pursue the problems without difficulty.

Finally, given that the plural problem (and its extensions) and the contraction problems would make up the linguistic part of the Unit, we suggest that the problems be ordered so that the plural comes first, then <u>want + to</u> contraction, and finally <u>x + is</u> contraction. We suggest starting with the plural because of the superficially less abstract nature of the phonology problem, the feeling it is somehow more physical and thus more readily available for discussion. The ordering of the contraction problems is based on our sense that the <u>wanna</u> problem is easier to solve than the <u>x + is</u> problem.

ORDERING THE CURRICULAR COMPONENTS

We expected in the pilot sessions to learn how to order the three curricular components. Unfortunately, given the piloting conditions we were unable to get as much information as we had hoped. Since the sessions were held after school, attendance was irregular and students tended not to complete the full set of lessons. In fact, in each group, there was only one student who attended all the sessions.

Judging, however, from the nature of the materials and student responses to them, we think it would be reasonable to order the components as follows: The King's Rule, the linguistics component, The Scientific Method, and the laboratory science component. In our next round of piloting, we plan to try this ordering.

COMPETENCY LEVELS OF STUDENTS

In our last technical report (STAMPS, January 1985), we raised questions about how students operate with data, questions that we wished to keep in mind as we analyzed the pilot sessions. Our interest is in whether students can solve the types of problems we present them with, whether they can organize data in a manner conducive to deriving rules and principles, whether they can formulate rules and apply them in a consistent enough manner to determine if they really work. We are also interested in differences among students in their ability to work with data- to derive and apply rules, and to look for principles and explanations.

We found that most students can solve the types of problems we presented, and that this warrants keeping these problems in our curricular Unit. The triads of The King's Rule are a limited and appropriate introduction to hypothesis formation and testing. The real-world problems offered in the linguistics component, for example, enable them to work in a domain of science where the material is familiar and accessible. Here they can generate data on their own and proceed to organize and describe it, again forming and testing hypotheses about the principles at work.

Students do vary considerably, we found, in their ability to formulate rules, and in the consistency with which they can apply them. In our previous technical report we asked whether there are levels of description of data which students achieve. We did find evidence of such levels, which we are calling "competency levels".

The competency levels are somewhat different from what we hypothesized at the start. All students were able to generate rules and form hypotheses that had predictive power. Students differed in what they did when faced with counterexamples, and in whether or not they sought counterexamples.

A counterexample is defined here as an example or piece of data that shows a rule to be wrongly formulated. It gives evidence that the rule needs to be revised. Such evidence can be of two kinds, depending on whether the piece of data is acceptable or not acceptable. Applying a misformulated rule will show either:

(1) The rule fails to predict an acceptable example; reformulate the hypothesis to include it; or

(2) The rule predicts an unacceptable example; reformulate the hypothesis to exclude it.

Thus, reformulation has to either include or exclude the new piece of data, depending on how the evidence makes the prediction fail.

In the piloting, students generally began by formulating a hypothesis, often on the basis of an inadequate sample, but a working hypothesis nevertheless, that had predictive power. Then, they started testing. They met a counterexample that showed their hypothesis to be wrong. What did they do?

Very different things. Some students were genuinely perplexed by data that their rule did not predict, and strongly resisted the idea that a set of data could be consistent with more than one hypothesis. Sometimes they reformulated their hypothesis to account for the last example only. Other times they decided the counterexample was a "mistake" and tried their original rule again and again despite repeated rejections of their predictions.

Other students realized that the new information needed to be integrated with the old. They reformulated their hypothesis, eventually achieving a successful revision, the number of passes depending on their accuracy and care in doing the revision.

Still other students anticipated examples that might not fit their hypothesis. This was the highest competency level: the students recognized that competing hypotheses were possible. They either formulated alternative hypotheses and sought additional data to decide among them, or they sought data to test and disconfirm their hypothesis. These students had no trouble, if they met counterexamples, in revising their hypothesis on the basis of all the data encountered.

What follows is a description of the four competency levels for hypothesis formation and revision that emerged from the piloting, with examples given for each one. The examples are drawn from the students' work with The King's Rule and with the linguistics problems. In all cases students began by formulating a workable hypothesis that had predictive power. The interesting variation occurred in their handling of counterexamples.

Competency Level 1: No Revision of Hypothesis

A. Consider the Last Example Only

When the students' hypothesis failed to predict an acceptable example, they considered only the newest piece of data, and formulated an independent hypothesis to fit this last case. An example of Level 1A from The King's Rule follows.

Example: First triad given by program was 10-9-5. The group formulated a rule to cover it: subtract 1 and then 4. They generated other triads with this rule; each one worked. They moved to the quiz. When given 12-10-6 in the quiz they said "no". When informed that the answer was "yes", they decided to try the rule subtract 2 and then 4, failing to keep in mind that their new rule should also cover 10-9-5. They found that triads generated with this new rule worked. To 8-7-5 in the next quiz they said "no", and again were informed they were wrong. They then tried the rule subtract 1 and then 2, once again considering the last example only.

B. Only One Rule Works

When the students' hypothesis predicted an unacceptable example, they needed to reformulate their hypothesis to exclude this example. However, they could not do so. They failed to recognize that the data could be consistent with more than one hypothesis and persisted with their original rule. An example of Level 1B from The King's Rule is given.

Example: First triad presented was 4-12-36. The group formulated the rule add 8 then add 24, and suggested the triad 3-11-35. When the program said "no", they said "But the first you go up 8 and the second you go up 24!" and tried again, suggesting the triad 6-14-38. Again they were told "no". "Why's it doing it wrong?" was one reaction. The students felt that their rule must work, since it worked once.

Students working at Level 1 were unable to synthesize new data with existing data.

Competency Level 2: Partial Revision of Hypothesis

At this level, students revised their hypothesis by including or excluding a new piece of data. They revised their hypothesis on the basis of all the data, but did so inadequately. The new rule failed to make good predictions, even though they tried to make it appropriately narrower or broader. An example of Level 2 is Group 3's formulation of the plural rule.

Example: In working through the plural rule, students had recognized that both /s/ and /z/ occur as forms of the plural. Their hypothesis at this point was simply:

(c') Add a /s/.

Add a /z/.

The instructor provided a counterexample, a case where the rule permits an unacceptable form: <u>house</u> --> *<u>house/z/</u>. One student revised hypothesis (c') to (c"):

(c") Add a /s/ or /z/ depending on how the word ends.

This rule was an improvement on the previous hypothesis because it indicates that the distribution of /s/ and /z/ follows from some fact about the difference in the way the words end. However, this rule still permits *<u>house/z/</u>, an unacceptable form. The next revision (d) was better since it included the full set of plural endings. However, the notion "how the word ends" still must be specified.

(d) Add a /s/ or /z/ or /iz/ depending on how the word ends.

Note the difference between this example and The King's Rule cases described in Competency Level 1. In those cases, rule inadequacy was distinguished in two ways: rules which failed to predict an acceptable example, and rules which predicted an unacceptable example. These two types of rule inadequacy, which are quite separate under the conditions of The King's Rule, fall together in the linguistics problems. The reason is that in the case of language, all the data are readily accessible.

The language example is one in which the rule predicts an unacceptable form, *<u>house/z/</u>. This is similar to Competency Level 1B, where the program informs students that what their rule predicts is not acceptable. But in this case, students have more information available -- namely, they already know what an acceptable form is. They are not simply stuck with just negative information about what must be excluded. They already know the acceptable form that their rule is not predicting; thus, they can reformulate their hypothesis to include it. This is a real benefit of using language as the domain for investigation.

Competency Level 3: Full Revision of Hypothesis

To include or exclude a new piece of data, students thought through the consequences of their hypothesis and revised it to acount for all the data collected so far. An example of Level 3 is Group 1's formulation of the plural rule.

> Example: In working through the plural rule, students had accounted for the endings /s/ and /z/. They recognized that the ending depended on the final sound of the word to which it was added, and also that the relevant factor was voicing. Their

hypothesis at this point was:

(d) Anything that makes your voicebox vibrate (like b,d,g) gets a z sound.

Anything that doesn't vibrate gets s.

When given the word thrush, they recognized its plural form as a counterexample. The correct form is not the one the rule predicts, *thrush/s/, but rather, thrush/iz/. As a step forward toward a revision they began by listing types that require /iz/:

(e) Any word with sh [e.g., thrush] takes ez.

(f) Any word with se [e.g., horse] takes ez.

(g) Any word with the s sound takes ez.

Some discussion of the types of sounds on this list revealed that they all were "hissing sounds", and the students revised their earlier hypothesis (d) to (h):

(h) Anything that makes your voicebox vibrate gets a z sound.

Anything that doesn't vibrate gets s.

But, if it's close to an s or a hissing sound it takes ez.

This was a correct revision of the hypothesis taking into account counterexamples, together with all the previous data.

Competency Level Four: Recognize Possibility of Competing Hypotheses

Students anticipated counterexamples that might not fit their hypothesis. They recognized that competing hypotheses were possible. They either formulated alternatives to their hypothesis and sought additional data to decide among the hypotheses, or they sought specific data to disconfirm their hypothesis. Examples of this level are taken from The King's Rule and Group 3's formulation of <u>x + is</u> contraction.

Example 1: The first triad offered by the program was 1-9-17. The students hypothesized <u>add 8</u>. "Wait," cautioned one student, "maybe it's <u>differences of 8</u>. Try going backwards: 17-9-1." Here the student recognized that the example was consistent with more than one hypothesis, and sought additional data to decide between two obvious possibilities.

Example 2: The students had made observations of the point of interpretation of wh-words in sentences for which <u>x + s</u> contraction was possible and those for which it was not. The students had arrived at the following hypothesis:

(c') Can only contract <u>is</u> if the wh-word is interpeted at the end of the sentence.

The instructor asked the group how they would test this hypothesis. The students discussed what would be a valid counterexample to falsify the hypothesis (thus, indicating there was another explanation for the data). Initially, one student incorrectly suggested that a counterexample would be a sentence where the wh-word was at the end of the sentence, but where contraction was allowed. Later, he decided a true counterexample would be a sentence where the wh-word was not at the end of the sentence but where contraction was allowed. Rule (c') was then tested by lengthening a version of one of the data sentences:

data: I wonder what Emma is looking at.

Emma's - ok

test: I wonder what Emma is giving to Peter today.

Emma's - also ok

The students realized this was a counterexample and proceeded to reformulate their hypothesis.

FURTHER PLANS

The curriculum component still to be added to our Unit will focus on a series of laboratory experiments. Here the students will deal with physical phenomena and carry out experiments in class to examine these phenomena. This is the part of the curriculum where students get direct experience with physical experimentation, and have a chance to practice research methods.

Thus, the work for the third year of the Scientific Theory and Method Project requires the design and piloting of additional curriculum materials, the revision of the piloted materials, the design of an evaluation instrument, the preparation of the Unit for classroom use, a month-long piloting of the entire Unit in two classrooms, and an evaluation of this month-long piloting.

APPENDIX A: PILOTING SCHEDULE

Group 1: 7 students, grades 9 and 10

Linguistics sessions:

- 1/16 plural formation
- 1/23 plural formation and its extensions
- 1/28 extensions of plural, cont'd.; past tense formation
- 2/4 tag question formation

Software sessions:

	subgroup 1		subgroup 2
2/13	The King's Rule	2/11	The King's Rule
2/25	The Scientific Method	2/26	The King's Rule, The Scientific Method
3/12	The Scientific Method	3/4	The Scientific Method

* * *

Group 2: 5 students, grades 9 and 10

Software sessions:

	subgroup 3		subgroup 4
1/14	The King's Rule	1/22	The King's Rule
1/28	The King's Rule	2/26	The Scientific Method
2/8	The Scientific Method	3/5	The Scientific Method
2/15	The Scientific Method		

Linguistic sessions:

- 3/18 plural formation
- 3/25 plural formation, cont'd.
- 4/1 x + is contraction; want + to contraction
- 4/10 yes/no question formation; negative sentence formation

* * *

Group 3: 7 students, grade 7

Software sessions:

	subgroup 5		subgroup 6
3/26	The King's Rule	3/28	The King's Rule, The Scientific Method
4/2	The Scientific Method	4/4	The Scientific Method
4/24	The Scientific Method	4/11	The Scientific Method

Linguistic sessions:

- 4/30 plural formation
- 5/14 want + to contraction
- 5/21 x + is contraction

APPENDIX B: EXCERPT FROM THE KING'S RULE TEACHER'S GUIDE

Each level of difficulty in The King's Rule is described below together with examples of characteristic rules. This material is excerpted with permission from The King's Rule: Mathematics and Discovery Teacher's Guide (O'Brien, 1984).

Level 1: The Castle Gate

Rules in Level 1 involve small upward or downward jumps, from 1 to 10 inclusive, from the first number to the second to the third, with all numbers lying between 1 and 30. All sets presented by the program in Level 1 are transparent. That is, they clearly suggest the rules that govern them.

Examples:	10-12-14	Increasing jumps of 2
	25-20-15	Decreasing jumps of 5

Level 2: The Guards' Room

Rules found in Level 2 involve constant multiplication or division jumps by numbers between 1 and 9 inclusive, using "seed" numbers between 1 and 5. The sets presented by the program in Level 2 are also transparent: they clearly represent the rules that govern them.

Examples:	2-10-50	Numbers multiplied by 5 - ascending
	48-12-3	Divide each succeeding number by 4

Level 3: The Game Parlour

The rules in Level 3 include all of those used in Levels 1 and 2, with the addition of ten new rules. Again, each set presented by the program clearly suggests its rule.

Examples:	Rule 1:	All numbers end in 9 , e.g.,9-29-19.
	Rule 5:	All numbers are divisible by 3, e.g., 12-6-18
	Rule 9:	All of the numbers are odd, e.g., 13-17-7

Level 4: The Magician's Study

Level 4 calls for deep thinking. Here the third number is the result of an operation -- addition, subtraction, multiplication, or division -- involving the first two numbers, and in some cases a constant. [18 rules result from formulas where:]

A [first number]: random value from 1 to 5
B [second number]: random value from 1 to 5
C: a value ranging from 1 to 5, staying constant throughout study of the rule
INT (Integer) means ignore fractional parts.
ABS (Absolute value) means ignore the negative sign.

Examples: Rule 4: [triad produced is] A-B-(B - A)
Rule 8: [triad produced is] A-B-(INT (A + B/2))
Rule 10: [triad produced is] A-B-(ABS (A - B))
Rule 14: [triad produced is] A-B-((A * B) - C))

Level 5: The Royal Suite

Level 5 differs significantly from previous levels. Here, the example sets given by the computer may be deceiving in regard to the rules they represent. Therefore, it is especially important to critically test several different hypotheses. If the student is presented with the set 100-200-300, the rule may not be the obvious ("multiples of 100" or "jumps of 100"), but something not so obvious such as "multiples of 2" or "multiples of 5".

Suppose the example set, 100-200-300, represents the latter rule "multiples of 5". Many players are likely to assume that the rule could be "multiples of 100"...such as 300-400-500. Since these numbers are also multiples of 5, the test sets would be approved. When presented with the quiz, players might then be asked to approve or reject the numbers 100-200-45. They might think "those numbers are not multiples of 100" and accordingly reject the set. Such an answer would be incorrect, and the players would fail the quiz. A better strategy would have been to test for the rules "multiples of 10", "multiples of 5", "multiples of 2", and others. In Level 5, students must be constantly on guard against being misled, although the program will always tell the truth.

Following are the guidelines for each "disguised" example given by the program, and the actual rule each represents:

1. Example: Three consecutive even numbers.
Rule: The second number is greater than the first, but less than the third.

8. Example: Three consecutive multiples of 10.
Rule: The sum of the numbers is divisible by 10 (e.g., 8-3-9).

11. Example: The second number equals the third. The first number is the second number times the third.
Rule: At least two of the three numbers are equal.

18. Example: The three numbers are all even.
Rule: The sum of the numbers is even.

Level 6: The King

Level 6 uses very simple rules, but is extremely challenging because the rules may apply to only one or two of the numbers presented by the program. For example, the program may present 64-66-68 and the rule would be "the second number is even".

Following are the guidelines for each "disguised" example given by the program, and the actual rule each represents:

1. Example: The first number is a multiple of 100 plus 1. The second number is the first plus 1, and the third is the second plus 1
   Rule: Only one number is even.

4. Example: The second number is 1. The third is the same as the first.
   Rule: The second number is a square.

11. Example: The third number is the square of the sum of the first two.
   Rule: The third number is the largest.

15. Example: Each of the three numbers has the form ABC,ABC (typical sets presented by the computer would be 947,947-345,345-169,169 or 12,012-321,321-2,002).
   Rule: At least one number is divisible by 13.

APPENDIX C: TRANSCRIPT FROM AN EXIT FOR THE MARBLE PROBLEM

At this point in the program, the students were asked which box they wanted to take a marble out of so as to correctly label each incorrectly labelled box. The following conversation between the instructor (I) and four students (A, B, C, D) occurred during an exit.

A: BW. [Says this right away, reaching for the keyboard.]

I: Wait. Before you go on, let me ask you which box you think you should take the marble out of.

A: Um...BB.

I: You would look at BB?

A: Well, no. BW.

I: You'd look at BW?

A: Yeah.

I: And you?

B: I'd say BB.

I: And you, C?

C: BW.

I: Okay, tell me why you would look at BW.

A: Well, actually I would look at BB, because if you picked BW you wouldn't be able to tell. But, if you chose BB you could tell.

I: Okay, someone else?

B: I'd say BB.

I: Why?

B: For the same reason. If you picked BW it could be either one.

I: What would you say, C?

C: I don't know. I think I would take BW because of the difference between the three boxes. And then you could tell the difference.

I: Okay, tell me a little more.

C: Because, you could tell the difference between the black and white marbles and then you could tell what was in the rest of the boxes.

I: Okay, so what do you think D? What would you go for?

D: BB.

I: Okay, I want you to think about it again. Let's say you take the cover off BB and there is a black marble there. I want you to tell me where the BB label goes?

[At this point, the students shrug their shoulders.]

A: You wouldn't really know.

I: Then why would you take the marble out of BB?

[The students shrug their shoulders.]

I: Now, with that idea in mind, thinking it through, which box would you choose to look at the marble from?

A: BW.

I: Why?

A: Well, there is no way you could find out what the other one would be except for if you did BW. ...Well, for all of them you wouldn't be able to tell.

I: Okay, now, what do you think, C?

C: BW.

I: And the reason is? Explain it to me. Give me your thought process.

A: Well, I don't think you would be able to find out whatever it [the other marble] would be because if you picked BW you could get a white and then you wouldn't be able to know if the other one was black or white.

I: Wait a minute. Let's think of the conditions here. You have three boxes. You know that these labels are incorrect. We know that there is a BW pair, a WW pair, and a BB pair. These labels are incorrect. We know that there is no way a BW pair could be in the BW box. So, if you draw a white marble out of there [BW box], then what would the other marble be?

A: It would have to be white.

I: Does that make sense?

B: No, because it could be black.

A: No, because...

I: If you pull a white marble out of here how could the other be black?

B: Oh. Okay, I get it.

The instructor then worked through the rest of the boxes in a similar manner showing that taking a marble from the other two boxes (BB or WW) would not be helpful in concluding which labels should go where. He did this until each student understood the reasoning behind the decision. He told the students, "So the purpose for all of this is just to say, if you have a problem you have to think it through."

APPENDIX D: TRANSCRIPT FROM AN EXPERIMENTAL METHOD SESSION

The following discussion occurred at the beginning of the second session of The Scientific Method with subgroup 5; three weeks had passed since the first session. An instructor (I) and two students (K, R) were present.

I: I want to ask you a couple of questions. First of all, can you tell me what an independent variable is?

K: It's the variable that...

I: Okay, how about control?

K: It's...you have to keep the controls the same to run an experiment.

I: Okay, so what do you think independent variable is?

K: The amount of something or what the amount is?

I: Okay, B, you want to take a guess at it?

R: No.

I: Okay, you guys remember the experiment with the water in the cans? You had three different metal containers and they each had water in them? Do you remember them?

K: Oh, yeah.

I: Do you remember what the controls were for that?

K: The temperature of the water.

R: And the amount of water.

K: And the time of boiling the water.

I: Okay, so what was the independent variable?

R: The amount of water.

K: No, that was a control. It was the same.

R: Oh, yeah.

I: Controls are the same, right?

K: Yeah.

I: Okay, so what is the independent variable?

R: The different kind of cans.

K: The kind of metal.

I: Why is that an independent variable:

R: Because they are different kinds of cans. They are not all the same.

I: So what is independent variable.

K: The different kinds of metal.

I: Right. But, in any experiment what is an independent variable?

K: Oh, uh...

R: Different things.

K: Like, okay, if you have an egg and then a yellow egg or if you have like a smaller thing and then a larger thing, things like that. I really don't know how to explain it.

I: Okay, let's try this. We are going to do an experiment. What is the first thing we want to do?

K: We want to do the experiment.

I: Okay, but how do we set up the variables?

R: We find the different things.

I: Yes, but, do we define the problem first? Show we have a hypothesis?

K: Yes. [Laughs.]

I: Okay, we make a hypothesis and we want to set up an experiment.

How do we set up the experiment so we can see if our guess is right? [Pause.] Well, remember the Bermuda grass? The student said I am going to test the power of love and he said the grass I love should grow a lot and the grass I don't love should not grow a lot.

K: Oh, yeah.

I: What was wrong with that experiment?

K: He put water on one grass and didn't on the other.

I: Well, what was wrong with that?

R: Well, one gets all the attention and gets water and the other didn't.

I: So what's wrong with that? Why isn't it a good experiment to tell whether it is love or not?

K: Because he didn't do the same things for both of them.

R: He didn't treat one like the other. He didn't treat the one with love like the other one.

I: And he should have?

K: Yeah.

I: So what does that mean? What are those things that are all the same? Those are...?

K: Controls.

I: And if you change one thing? That is...?

K: The independent variable.

I: Okay, so do you have a better grasp of that?

K: Yeah.

I: Okay, so in the water experiment, all those things that were the same are...?

K: The independent variable was the kind of metal. All those that were the same were the controls.

I: Okay, so if you were going to set up an experiment would you change a lot of different things?

K: No, you wouldn't really be sure of how it turns out.

I: Okay, so how would you set up the experiment using the words independent variable and controls?

R: You would test them all the same and some not the same.

I: Some not the same?

R: Yeah.

I: Okay, so you would have some things the same and some things different?

K: Yeah.

I: Okay, what if I said with the Bermuda grass that the amount of sunlight is the same, the amount of time with the sunlight is the same, and that all ! am changing is the amount of water and the amount of love. What is wrong with that?

K: You are not giving them the same amount of water?

R: It has to be the same amount of everything. If you give one too much water you have to give the other too much water.

I: But you just told me that some things can stay the same and that some things can change. How many things can change?

K: One.

I: Okay, that's the idea with independent variable.

R: Like if you plant a see, two seeds, you have to keep them both in the same amount of dirt, you have to keep them in the same amount of deepness, and the same amount of sunlight and give them the same amount of water.

I: And what would be the one things you would change?

K: Amount of love.

I: Okay, what if I said I want to test whether water affects the growth of grass or not.

K: You would give one water and one not.

I: And what about the other stuff?

K: You would keep everything the same.

R: Yes, everything the same except the amount of water.

APPENDIX E: EXCERPT FROM "NAVAJO LINGUISTICS" (HALE, 1975)

The following material is excerpted with the author's permission.

One might well ask why it is useful to study a language which one already knows. The answer to this question is a rather long story, but it is the same as the answer to the question "why do we study biology, chemistry, and physics?" "Why do we study science at all?" The reason is that we wish to find explanations for the things that we observe. We observe, for example, that water freezes at a certain temperature, and it boils at another temperature. Gasoline is also a liquid -- it flows like water -- but it doesn't freeze at the same temperature as water. Also, unlike water, it burns if you touch a lighted match to it. We are not satisfied with just noting these facts; we want to know why they act differently. Whenever we look beyond the things we observe and try to explain them, we are engaged in science. The scientist looks deeper into the structures of water and gasoline in order to discover what properties they have that make them behave the way they do.

The study of language -- i.e., linguistics -- is also a science. We know that people are able to speak languages, but we know very little about what that really means -- a person knows his native language (and some people know several languages), but what exactly is that knowledge like? We observe that a person can understand the sentences of his language and that he can speak the sentences of his language. The question is: why is he able to do this? The linguist tries to answer this question. He tries to construct a theory which will account for this ability.

Now, in any science, no matter how concrete it is, we are involved in theory-building -- i.e., in the formulation of statements which will best account for the facts we observe. In fact, we are involved in this activity all the time, whether we call ourselves scientists or not. Everything we feel relatively sure is true of the world is a theory. If we are told, for example, that there is a crow sitting on a fencepost outside, and if we have seen crows before, we do not wonder what color the crow on the fencepost is. We predict that it is black -- because we have a theory that all crows are black. But it is only a theory; we haven't seen all of the crows in the world, nor have we seen all the crows that went before and all of those as yet unborn. We have developed a theory that all crows are black on the basis of our observations -- i.e., all the crows we have ever seen have been black. On the basis of this theory, we are relatively sure the crow on the fencepost will be black.

The theory about crows is not a very startling one, to be sure. It is so commonplace that it is uninteresting. Nonetheless, it is like a scientific theory, in that it permits us to make predictions. Theories become interesting to us intellectually when they are less obvious. Thus, when we try to explain <u>why</u> all crows are black we are forced to construct a theory which is more abstract. One such theory is that there are genetic laws which determine the physical characteristics of animals -- it is part of the genetic make-up of crows that their feathers are black. This theory is more interesting because it attempts to explain much more than the color of crows; it explains why the young of any animal resembles its parents. Recent work in the field of genetics has added a great deal of detail to this theory and has actually isolated the material in which the genetic code is carried. When the theory was first proposed, however, it was highly abstract -- it was the best explanation that could be suggested to explain the observation that certain physical characteristics are transmitted from parent to offspring.

I have strayed some distance from the topic of linguistics and the study of language. I have done so merely to point out that any serious science is concerned with theories. Linguistics is not a physical or biological science; rather, it is the study of a certain aspect of the human mind. We know that a person's knowledge of his language is stored in his brain, but we cannot observe it directly. What we do observe is his speech -- on the basis of this, we try to construct a theory of what is in the brain. This is exactly what is done in other sciences -- if some object is not directly observable, a theory, or model, is constructed which can duplicate the observable behavior of the object. The theory is correct to the extent that it can accurately duplicate this observable behavior.

The linguist is in one respect better situated than other scientists. He does not need a lot of equipment to observe the data he studies -- he has in his head a knowledge of his own language; he can therefore observe his own speech. ... If done properly, such an endeavor will serve as a means of introducing the scientific method to students and will, thereby, contribute significantly to their educational development. It has the advantage over the other sciences that it makes use of material which is thoroughly familiar to the students -- i.e., their own speech behavior.

REFERENCES

Hale, Kenneth. (1975) "Navajo Linguistics", unpublished manuscript, Massachusetts Institute of Technology.

O'Brien, Thomas C. (1984) The King's Rule: Mathematics and Discovery Teacher's Guide, Sunburst Communications, Inc., Pleasantville, NY.

Scientific Theory and Method Project (STAMPS). (January 1985) Doing Science: Constructing Scientific Theories as an Introduction to Scientific Method. Unpublished Technical Report, Educational Technology Center, Harvard Graduate School of Education, Cambridge, MA.