DOCUMENT RESUME

ED 342 630 SE 052 258

AUTHOR	Glaser, Robert; And Others
TITLE	Scientific Reasoning across Different Domains.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
PUB DATE	91
NOTE	25p.
AVAILABLE FROM	Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Assisted Instruction; *Discovery Learning; *Heuristics; Higher Education; Independent Study; *Learning Strategies; Pretests Posttests; Science Education; *Thinking Skills
IDENTIFIERS	Microcomputer Based Laboratories

ABSTRACT

This study seeks to establish which scientific reasoning skills are primarily domain-general and which appear to be domain-specific. The subjects, 12 university undergraduates, each participated in self-directed experimentation with three different content domains. The experimentation contexts were computer-based laboratories in d.c. circuits (Voltaville), microeconomics (Smithtown), and the refraction of light (Refract). Subjects spent three 1.5 hour sessions working with each laboratory and took pretests and posttests that assessed their learning. Specific patterns of strategies used in each laboratory depended primarily on the structural form of the discovery task and the nature of the domain. In a situation that required the discovery of correlational regularities, evidence-generation activities, like the heuristic of controlling variables, were primary. In contexts where the regularities were functional rules, evidence interpretation became important. When the rules were quantitative, mathematical and algebraic heuristics were important. Students appeared very sensitive to the task demands of each laboratory, and adjusted their strategies accordingly. Regardless, they learned more as they proceeded from domain to domain, indicating that they were becoming more effective in planning and carrying out experiments, and in formulating and testing hypotheses based on those experiments. The findings suggest that the most generally useful skills for direct instruction may be those for evaluating the kind of problem at hand and for selecting the most appropriate processes and strategies. (Author)

SE

Scientific Reasoning Across Different Domains

Robert Glaser
Leona Schauble
Kalyani Raghavan
Colleen Zeitz

Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260

Center for the Study of Learning

LEARNING RESEARCH AND DEVELOPMENT CENTER

University of Pittsburgh

Abstract

This study seeks to establish which scientific reasoning skills are primarily domain-general and which appear to be domain-specific. The subjects, 12 university undergraduates, each participated in self-directed experimentation with three different content domains. The experimentation contexts were computer-based laboratories in d.c. circuits (*Voltaville*), microeconomics (*Smithtown*), and the refraction of light (*Refract*). Subjects spent three 1-1/2 hr sessions working with each laboratory and took pretests and posttests that assessed their learning. Specific patterns of strategies used in each laboratory depended primarily on the structural form of the discovery task and the nature of the domain. In a situation that required the discovery of correlational regularities, evidence-generation activities, like the heuristic of controlling variables, were primary. In contexts where the regularities were functional rules, evidence interpretation became important. When the rules were quantitative, mathematical and algebraic heuristics were important. Students appeared very sensitive to the task demands of each laboratory, and adjusted their strategies accordingly. Regardless, they learned more as they proceeded from domain to domain, indicating that they were becoming more effective in planning and carrying out experiments, and in formulating and testing hypotheses based on those experiments. The findings suggest that the most generally useful skills for direct instruction may be those for evaluating the kind of problem at hand and for selecting the most appropriate processes and strategies.

Previous work in scientific reasoning, our own (Shute, Glaser, & Raghavan, 1989; Schauble, Glaser, Raghavan, & Reiner, 1990) as well as others' (Langley, Simon, Bradshaw, & Zytgow, 1987; Klahr & Dunbar, 1988) has empirically investigated scientific reasoning in various discovery tasks with the objective of characterizing the strategic or reasoning processes associated with successful discovery of lawful regularities. Most of these studies have been carried out in the context of one domain of knowledge. However, we have noted as we work in different domains that there appear to be strong influences of the structure and content of the domain on the particular reasoning and inference skills that subjects employ. This observation has led us to investigate the reasoning of subjects who work to discover the principles that apply in three computer laboratories incorporating simulations of different content domains in the physical and social sciences.

Historically, most of the psychological research on scientific discovery has regarded scientific reasoning in one of two ways. Some studies investigate reasoning processes, in particular, strategies of scientific experimentation, such as designing and interpreting valid experiments, hypothesis testing, identifying regularities in patterns of data, and reasoning about correlation and covariation in events. This tradition tends to cast these skills as being rather general reasoning abilities that presumably are applied across content domains. Other work emphasizes the content and structural characteristics of domain knowledge as a function of prior misconceptions or as a function of expertise. Within this line of work, the emphasis is on strategies and heuristics that are quite specific to the domain and the task. When an individual is perceptive of the features of a problem, these heuristics often become proceduralized, with the consequence that they may be employed almost automatically when particular task requirements elicit them. For example, experts appear to solve physics problems by spontaneously perceiving and classifing the problems in terms of the underlying domain principles that comprise their deep structure, in contrast to novices, who focus upon the surface structure (Chi, Feltovich, & Glaser, 1981).

Empirical research on scientific reasoning is increasingly attending to the relations between domain-general strategies and domain-specific reasoning heuristics. For example, Kulkarni and Simon (1988) have reconstructed the reasoning processes employed by Hans Krebs as he solved a particular problem solved in the history of science, the discovery of the urea cycle. They concluded that some of the heuristics he employed were closely tied to the domain of biochemistry, whereas others were more general strategies applicable to discovery in all domains of science or to other forms of problem solving.

This study continues the investigation of the relations between general and specific reasoning in science. Kulkarni and Simon's conclusions were based on a reconstruction from historical records, such as Krebs' notebooks; we here move on to investigating these issues experimentally. Unlike Krebs, our subjects are university undergraduates who are novices in the domains of investigation. Each of our subjects participates in self-directed exploration in three different content domains, providing us with the opportunity to investigate which reasoning and inference activities are employed with some consistency and systematicity from domain to domain, and which activities appear to be used more narrowly within a more limited range of content.

The three computer laboratories used in this study simultate phenomena in the domains of economics (Smithtown), d.c. electric circuits (Voltaville), and the refraction of light through lenses (Refract). In each laboratory, students can construct experiments by varying variables and parameters, take relevant measurements, make predictions about outcomes, record and manage data, and develop and revise hypotheses about the laws and principles that apply in the domain.

METHOD

Subjects

Participants were recruited on a university campus. Since the study required relative novices in the domains of interest, criteria for acceptance in the study were that the candidate be an undergraduate majoring in a nonscience discipline. The first twelve applicants who fit these criteria were admitted as subjects, yielding a group of 4 men and 8 women, mean age 21 years (range from 18-25). No participant was currently studying physics or economics.

Procedure

Sequence of Experimental Sessions

The study was described to subjects as a study concerning learning with computer laboratories. They were told that they would be shown how to use the laboratories and then would spend several sessions working with each lab "as a scientist might" to try to discover as many laws and regularities in the domain as possible.

All participants took a brief test designed to screen for competence in simple algebra and in the ability to make qualitative and quantitative interpretations based on tables of numerical data. Subsequently, each subject came to the university laboratory from two to three times per week to participate in a total of eleven experimental sessions lasting one and one-half hrs each. Total duration of the study was therefore approximately 16 hrs for each subject, extending over six weeks.

Subjects were randomly assigned to one of two treatment orders. Because of the time-intensive nature of the study, a completely counterbalanced design was not feasible. Our task analysis predicted that Voltaville and Smithtown share the least amount of overlap in the activities and skills required for successful learning. In contrast, Voltaville and Refract overlap somewhat in their requirements for interpreting evidence, whereas Smithtown and Refract appears to require some common skills in generating evidence. Because Refract has mixed characteristics, sharing some task requirements with Smithtown and others with Voltaville, it was the most useful laboratory for studying consistency or transfer of reasoning from the other two labs. Six of the subjects worked for several sessions on Voltaville, and then on Smithtown, whereas for the remaining six, the order was reversed. All subjects worked last on Refract.

Working With the Laboratories

Work with each laboratory was preceded by a short pretest (about 20 min) to assess subjects' prior knowledge in the domain. Each pretest included qualitative questions addressing conceptual understanding. In addition, for those domains in which the relations take the form of mathematical expressions (Refract and Voltaville), pretests also included items designed to assess knowledge of and ability to apply these laws. After the pretest, an interviewer prompted the subject through a standard training and demonstration session (of about 40 min duration) with the appropriate computer laboratory. The purpose was to ensure that the subject understood the activities supported by the laboratory, could operate the computer interface, and was familiar with the discovery tools common to all three laboratories. After this demonstration was completed, the experimenter informed the subject of the task objective: to discover as many laws and regularities as possible. The subject spent the remainder of this introductory session in self-directed experimentation with the computer laboratory. In subsequent sessions subjects continued their exploration. Since the computer laboratories saved each student's activity to a personal file, experiments and records were preserved from session to session, and subjects started off each session with the information and discoveries they had generated in previous sessions. Thus, the study focused on learning that was cumulative over several sessions. In addition, since the computer records contained a complete trace of all student actions with the laboratories, they were a primary data source for the study.

During the learning sessions subjects worked individually with one of three interviewers. The interviewer answered questions about operating the laboratory but avoided directing student exploration. In addition, when appropriate, she prompted subjects to describe what they were thinking, to justify conclusions, and to explain what they were inspecting on the screen. These comments were recorded on audiotape.

Including the introductory sessions, each subject spent a total of three sessions working with Voltaville and Refract. Smithtown encompasses a somewhat larger domain, includes a greater number of goals to discover, and requires a greater number of experiments to support each hypothesis. Consequently, each subject spent four sessions working with Smithtown. At the end of the final session on each laboratory, students took a posttest composed of items parallel to the pretest items.

Analysis of Similarities and Differences Among the Domains

The three computer laboratories share a common interface and an identical set of tools that support the recording, sorting, and graphing of data, and the development of hypotheses. In addition to the laboratories sharing many common components and operating in the same manner, at the top level, the task posed to students working in each laboratory was identical: to try to find as many laws and regularities as possible. To discover the laws in these three laboratories, one must generate valid and informative experiments, record and manage the data from observations, and then appropriately interpret the data by developing generalizable laws. We refer to these classes of activities as the generation of evidence, data management, and evidence interpretation, respectively. However, because of differences in the overall structure of the domains, the experimentation strategies and activities that are most adaptive should differ from laboratory to laboratory.

The structure underlying Smithtown, the laboratory in microeconomics, is a correlational structure. Changes in certain dependent variables covary with changes to independent variables and parameters. The laws describing these correlational relations are qualitative statements of the form, "As price of tea increases, quantity demanded decreases," a principle in microeconomics known as the Law of Demand. Finding these principles involves generating evidence that supports appropriate inferences of inclusion (that is, identifying which variables are involved in a particular relationship, as well as the general direction of the relationship) and also exclusion (noting that some variables are not relevant in a particular relation and can therefore be ruled out of further consideration). Identifying these correlational relationships requires the generation of carefully structured patterns of evidence in which extraneous variation is controlled, and which thus permit the isolation of pertinent causal effects from other candidate causes. It is particularly important to avoid errors of false inclusion, that is, inferring that a variable plays a causal role when in fact other variables are also varying and therefore may be responsible for or contributing to the outcome. Therefore, it is likely that strategies and activities in the generation of evidence will be particularly important to successful learning with Smithtown.

In contrast, Voltaville, the d.c. circuit laboratory, is analagous to the classic rule discovery tasks widely explored in cognitive psychology. Pertinent examples include cryptogram tasks (Simon & Kotovsky, 1963) and Wason's (1960) 2-4-6 task. That is, the objective is to find a rule that correctly and exactly specifies the relations among all variables in the task. These rules take on forms such as "V = I times R." or "$R_1 + R_2 + R_3$ = total R." In such a rule discovery task the important operations are not inclusion and exclusion of relevant variables, but confirmation and disconfirmation of candidate rules where the relevant variables are apparent. Unlike the case with Smithtown, finding principles does not depend upon setting up carefully designed sets of observations that vary in prescribed manners. Rather, in Voltaville, each experimental observation is fully informative, since each observation embodies the laws that apply in a particular kind of circuit. For example, on the basis of measurements of the values in a series circuit with three resistors, it is possible for a subject to induce Ohm's Law, as well as Kirchhoff's Laws for Resistance, Current, and Voltage. Thus, it is likely that evidence generation strategies will be less important in Voltaville than in Smithtown. Instead, evidence interpretation skills are fundamental, including the use of mathematical heuristics.

Refract represents a mixed case. It is also a rule discovery task, with laws taking the form of mathematical expressions. A look at the handout indicates that the rules in Refract require more sophisticated mathematical knowledge than those in Voltaville, and strategies for the interpretation of evidence are likely to be important. However, as in Smithtown, one of the challenges in Refract is to identify the particular variables that are implicated when an independent variable or a parameter is manipulated. Managing the complexity of data in this laboratory is greatly facilitated if one systematically generates evidence in regular patterns. Since not all variables play a role in all laws, systematicity is particularly important in discerning which independent variables are responsible for changes in the corresponding dependent variables. Thus, strategies in the generation of evidence should also play an important role.

Beyond these differences in the structure of the three domains, there are important differences in the salience of the structure. The parameters in Refract and Voltaville represent changes in physical objects which can actually be manipulated: lenses made of different materials and shapes, circuits with resistors wired in series or in parallel. Subjects find it intuitively reasonable that changes in these parameters may change the way the entire system works. In contrast, the parameters in Smithtown are not easy to distinguish from the variables. Income level, interest rates, and price of a good all seem comparable, and subjects expect that they all have similar effects. Discerning the underlying structure of Smithtown is thus more difficult for most individuals.

In sum, Smithtown has a correlational structure, and the distinction between variables and parameters is particularly difficult to make in this laboratory. Voltaville is a rule discovery task, and the distinction between variables and parameters seems consistent with differences in the physical materials represented in the laboratory. Refract has a mixed structure. Since it is necessary to find out which independent and dependent variables are lawfully related, some correlational reasoning is required. On the other hand, the basic structure of the task is a rule discovery structure, with the objective of finding a rule that expresses the relations among the relevant variables and parameters. As in Voltaville, differentiating between variables and parameters is facilitated by the fact that parameter changes map onto changes in concrete physical materials like lens shape and material. These differential task characteristics should affect the use and character of exploratory and inference activities in these three laboratories.

Results

We first report student performance in the three laboratories, in particular, with respect to the task characteristics of each laboratory. Next, we discuss the extent to which subjects who work over an extended period with these laboratories both learn content knowledge and acquire proficiency in the processes of inference and discovery that lead to learning.

Experimentation Activity

Evidence Generation

First we consider the generation of evidence in the three laboratories. Generation of evidence encompasses the amount and breadth of search, the informativeness of search, and the structure of search.

Amount and Breadth of Search. The problem space comprising the number of possible experiments in scientific domains, sometimes referred to as the e-space (Klahr & Dunbar, 1988), can be very large. Furthermore, the informativeness of experiments designed will vary, with some regions of the e-space representing experiments that do not distinguish between rival hypotheses, and other regions representing comparisons that support definitive judgments about a hypothesis. The computer laboratories studied here have e-spaces that are quite large in comparison to those employed in many laboratory tasks. Of the three,Voltaville supports the smallest e-space: it includes three major variables (voltage, with 40 possible values, resistance, with 10, and current, a dependent variable that varies as a function of the values of the other two), and one parameter (circuit type) with eight different levels. In contrast, Refract has two variables (image distance, with 5 values and angle of incident ray, with 7) and two parameters, the relative optical density of lenses, with four levels, and lens shape, with eight levels. In contrast, Smithtown has only one variable, price. However, this variable has an exceptionally large range of values, since it is possible to vary dollar costs in various markets. In addition, Smithtown includes eight parameters (such as income level, population, interest rates, weather, and the like) which also have a very wide range of permissible values that shift the relations among the simple variables. Most subjects find it more difficult to identify the way that parameters work than to discover lawful variable changes (Shute et al., 1989; Schauble et al., 1990). Therefore, the relative proportion of parameters and variables, as well as its overall larger e-space, make Smithtown the most complex and difficult to master of the computer laboratories. For the same reasons, Refract is of intermediate complexity, and Voltaville contains the least complexity, both in amount and kind of possible variation.

As Table 1 shows, the larger the e-space supported by a lab, the more experiments subjects actually generated. Thus, our subjects appeared to be sensitive to the conditions under which more variation is possible, and responded by searching more broadly, thus generating more information. In addition, Table 1 shows that on the average, students made more changes to parameters in Smithtown than in either Voltaville or Refract, and changed variables more frequently in Refract and Smithtown than in Voltaville, a straightforward reflection of the differences in domain structure.

Insert Table 1 About Here

Informativeness of Search. Although each of the computer laboratories permits the generation of many potential experimental combinations, there is for each a much more tractable number that comprises the minimum set required to discover all the laws. This minimal amount of evidence varies from a low of only 6 experiments in Voltaville to 20 in Refract and approximately 50 in Smithtown (the number fluctuates somewhat depending on the path of experimentation). Consequently, not only does Smithtown have the largest and most complex e-space whereas Voltaville has the least; in addition, the minimal amount of evidence that must be generated to discover all the laws and relations is also greatest for Smithtown and least for Voltaville. As Table 1 shows, subjects typically generate smaller percentages of the minimum required evidence in Smithtown and Refract, a reflection of the larger and more complex evidence patterns required in those laboratories. On the average, students generate all or nearly all of the evidence required to support discovery of all eight laws in Voltaville, even though they may not go on to infer them. In contrast, they generate only half the evidence required for discovering Smithtown's twelve laws.

Structure of Search. Although subjects may operate in the most informative regions of information, they may still fail to structure their experiments so that they support valid inferences.

As discussed above, in Smithtown, laws are qualitative relations, whereas in Voltaville and Refract, they are mathematical expressions. Furthermore, for any one law being explored, most of the factors in Smithtown do not play a causal role, whereas in Refract and Voltaville, all the factors are indeterdependent. Because of these domain differences, discovering the laws in these three worlds entails structuring experiments in different ways.

To discover a law in Smithtown, students must generate three price points at several levels of a relevant parameter. In contrast, in Refract, relevant comparisons are pairs of observations that differ by only one variable change. This experimentation pattern is less complex than the structure of informative experiments in Smithtown, and there are more alternative paths to solution. In both Refract and Smithtown, conclusions are based upon noting regularities in changes from one observation to the next. If the comparisons are not valid, no definitive conclusion can be drawn. In contrast, in Voltaville all observations include information that can support valid inference. To yield meaningful data, there is no need as in the other laboratories to design a set of coordinated experiments that serve as contrasts, because each observation stands alone in supporting the induction of the relevant laws.

To generate valid patterns of evidence, it is necessary in Smithtown and desirable in Refract to follow the pattern of varying only one variable at a time, holding all other variables constant. As indicated, experiments in Voltaville are informative whether one varies one variable, two variables, or many. Our subjects appeared to be aware of this task structure. As Table 1 shows, the percentage of experiments in which subjects controlled extraneous variation was very high in Smithtown, and only slightly lower in Refract, but much lower in Voltaville. Note that although students generated controlled experiments much less frequently in Voltaville than in the other labs, they still did so nearly one third of the time, a substantial use of an evidence-generation strategy, given that there is no discernible advantage to using it here. Perhaps this performance reflects the fact that control of variables is one of the most commonly taught strategies in science instruction.

Evidence Interpretation

Generating valid and informative experiments is a necessary but not sufficient condition for discovering the laws in the computer laboratories, for obviously it is also necessary to appropriately interpret and make inferences about the evidence generated. We turn next to strategies in evidence interpretation that can be identified in the three computer laboratories.

Making Predictions. Predictions serve both as the products of inferences and as the engines for further inference. As Table 1 indicates, students more regularly made predictions about the outcomes of their experiments in both Voltaville and Refract than in Smithtown. We observed from protocols that subjects appeared to find it much more satisfying to generate a specific quantitative prediction, which was then unambiguously confirmed or disconfirmed by the computer feedback, rather than to generate a qualitative prediction such as, "Quantity demanded will decrease." Confirmation and disconfirmation of qualitative predictions of this kind are seemingly more ambiguous, and students appeared to find the feedback less helpful or satisfying, apparently because a mere correlational statement provides less information. The more informative feedback apparently results in more hypothesis-driven search. When a subject's prediction is disconfirmed in Voltaville or Refract, he or she learns not only that the working hypothesis is wrong, but specific information about how it is wrong, information which can be used in revising the hypothesis or generating additional informative search. In Smithtown, subjects in the same position learn only that they are wrong, with no special constraints to guide further search except the information that this particular statement should be eliminated as a hypothesis. Despite this point, attempts at inference through predictions resulted in an equal percentage of correct predictions in all three laboratories, averaging about three quarters of the time.

Prior Knowledge. Evidence interpretation is also influenced by prior knowledge. In general, subjects have experience with buying and selling, and therefore have a great deal of knowledge about consumer and market behavior. As a consequence, they hold a number of expectations about likely causes and effects in Smithtown, which might be correct or false misconceptions. Most of our subjects reported that they were much less knowledgeable about the physics domains, in particular, stating that they knew more about economics than electricity, and more about electricity than refraction. These differences in prior knowledge may either help or mislead subjects in deciding where to search for relations, may influence them to be more or less active in search for disconfirming evidence, and may affect their confidence in their conclusions as well as their ability to remember and apply the laws they discover. Differing prior knowledge could also affect the tendency to check to see whether a candidate law makes "sense" consistent with one's understanding of the phenomenon being described.

Table 1 shows that subjects stated a greater number of alternative hypotheses of all kinds (general and specific, correct and incorrect) while working with Smithtown, in comparison to both Voltaville and Refract. Most of our subjects were not hesitant to try out these tentative conclusions by submitting them to computer evaluation, even if little relevant evidence was available. However, this prior knowledge was a mixed blessing. On the average, subjects discovered a smaller percentage of the Smithtown goals than in either Refract or Voltaville. The mean percentage of goals discovered was 52.1%, 58.3%, and 88.5% in Smithtown, Refract, and Voltaville, respectively. Prior knowledge sometimes helps subjects to interpret patterns of evidence, but if prior knowledge is incorrect or only partly correct, it can encourage subjects to distort, ignore, or selectively interpret the evidence that they generate. This finding is a common one in research on scientific reasoning.

Data Management

The differences among the laboratories also result in differences in how students manage their memory by recording and organizing data. In Smithtown, laws often involve parameter changes that result in function shifts. Consistent with this characteristic, we found that our subjects graphed data more frequently in Smithtown than in the other two laboratories. As mentioned, the Refract laws are moderately complex mathematical expressions. As Table 1 shows, in Refract students were particularly likely to use the computer capability for organizing tables, with its spreadsheet sorting and expression-generating functions. With the exception of one relation, the laws in Voltaville are algebraically simple, and thus there was less necessity for students to store or organize data to support the discovery of the relevant laws.

In sum, differences in domain content and structure were associated with differences in task requirements from laboratory to laboratory. These, in turn, were associated with different patterns of student activity as they detect domain differences in self-directed exploration. The results just described are corroborated by a pattern of intercorrelations run across the relevant activities for all twelve subjects. These correlations reflected no activities in which subject performance was higly related across all three laboratories. Where strong correlations did exist, they were between pairs of laboratories, and they reflected the general structural and task differences already discussed.

In sum, there is no simple story about consistency of performance, at least at the group level. In general, our students did not tend to apply certain activities and processes across the three domains. Instead, the general picture is one of adaptiveness to the constraints of the task at hand. Those relations that did appear, were located in the discovery components in which laboratories shared common structural or task requirements.

Learning and Transfer in the Computer Laboratories

What does this pattern of specificity of performance imply for student learning? At least at the top level, the tasks posed by all three computer laboratories are the same. Students generate experiments, take measurements, make predictions, record data, and develop and revise hypotheses about the laws that apply. Much work on scientific discovery proceeds from the assumption that subjects differ in their skills or abilities to perform these activities. Our own earlier work proceeded from similar assumptions toward the objective of identifying patterns of activity that account for effective and ineffective learning (Schauble, Glaser, Raghavan, & Reiner, 1990; Shute, Glaser, & Raghavan, 1989). However, it appears based on our current results that these patterns are seriously attenuated by domain characteristics. These specificities in performance have two kinds of implications for student learning. The first concerns the way that students employ the skills they have to learn specific content knowledge. If individuals have differential skills, it is to be expected that some will do better in some domains, and others will do better in others. The second implication is for students' growing capability in learning how to learn. A group of students who possess the identical skills in scientific reasoning may still vary considerably in their self-regulatory skills for determining whether their skills are appropriate in a particular circumstance, how those skills will be applied, and what other skills need to be developed. Those of our students who learn effectively in more than one domain succeed not by generally applying an invariant set of skills, but by reacting more adaptively than other students to the fluctuating task demands posed by the three laboratories.

To explore these issues, we measured amount of learning for each computer laboratory by computing student pre/post test gain scores. Students accomplished significant gains in each of the laboratories. Their gains were relatively higher in Voltaville and Smithtown than in Refract, the most difficult discovery context. Mean gain score for Smithtown was 26.5 percentage points, for Voltaville was 26.3 percentage points, and for Refract was 11.9 percentage points (all of these gains are significant).

However, there was no clear relationship between amount of achievement in one laboratory and amount of achievement in the others. A correlation run on the three gain scores for the 12 subjects yielded only a modest correlation between gains in Refract and Voltaville ($r = .27$), the two worlds in which the rule discovery structure was shared. There was a *negative* correlation between gains in Voltaville and Smithtown ($r = -.31$), the two labs with different structures, rule discovery versus correlational. There was no meaningful correlation between gain scores on Smithtown and Refract ($r = .008$), which has mixed properties, so that students who were effective in Smithtown varied in the performance in Refract, and vice versa. In general, then, at the group level, learning in our laboratories appeared to depend to a large extent upon adaptibility to structural and task requirements rather than the exercise of generalized reasoning strategies.

This specificity of student performance is also manifested by the fact that in these complex exploratory situations, no student was clearly the best in all laboratories and no student was the worst. No student was among the one-third with the highest gain scores on all three laboratories (although five students were among the top third achievers on at least two labs). Similarly, no student was among the one-third with the lowest gain scores on all three laboratories (although three students were low achievers on at least two labs).

Thus, some students were better at some kinds of learning than others. But are students really this specific in their strengths?[1] We have empirical data at the group level which imply some general characteristics of performance. This general character may lie not only in the ability to adaptively apply relevant skills, but also in the ability to evaluate them. This is indicated by our data, which show that the subjects showed more and more learning as they progressed over the three different lab experiences. On the average, there was a mean increase in gain score of 9 percentage points from Lab 1 to Lab 2, regardless of whether subjects began with Voltaville or Smithtown (to evaluate the magnitude of this increase, recall that total gain score for each of these labs was about 25 percentage points). Thus, not only did the students adaptively apply their skills, but at a more general level of understanding, they became more familiar with the overall activity of experimentation and its component processes, including ways of generating evidence, making inferences from this information, searching for regularities, and testing them.

Furthermore, generality of ability in learning how to learn is revealed by a comparison of student learning in Voltaville and Smithtown, the first two labs, with their learning in Refract, the final laboratory explored by all students. From the group of twelve subjects, four were identified, for want of a better word, as "improvers." These were the students who made the greatest increase in learning gains when their gain scores on the second laboratory were compared to their gain scores on the first laboratory. Average increase in gain scores among this "improver" group was 33.3 percentage points from the first laboratory to the second. A second group of four made the smallest increase in gain scores, an average increase of -9.5 percentage points. On the third lab, Refract, the improvers gained an average of 19.5 percentage points from pretest to posttest. In fact, three of the four improvers made the largest overall learning gains in Refract, as well. In contrast, the non-improvers' mean gain score was only 5.8 percentage points, indicating that the amount of improvement from Lab 1 to Lab 2 was associated with the amount of learning achieved in Lab 3. Apparently, those students who learned the most about the general objectives and nature of scientific discovery by working with the earlier two labs were able to apply this understanding in Refract. In addition, since Refract is a lab with a mixed structure, it represented an opportunity for subjects to apply the particular relevant skills practiced in both Voltaville and Smithtown.

In summary, then, although for individual students low or high learning in one laboratory was not directly associated with low or high learning in the others, on the average for the group as a whole, students appeared to learn how to learn with computer laboratories. This appeared to involve becoming more sensitive to task similarities and differences from domain to domain, and learning how to adapt their experimentation activities accordingly.

Discussion

Recent research on experimentation has increasingly addressed the complexity of scientific discovery by studying the entire cycle of planning, designing, carrying out, and interpreting experiments, in contrast to earlier work, which typically focused on one of these component processes at a time, such as how people interpret disconfirming evidence. What contribution is being made by studying larger, more coherent episodes of scientific reasoning? One robust conclusion, consistent with our findings here, is that experimentation involves a complex orchestration of activities, and there is typically a great deal of variability in people's performance on the component processes. There are many alternative ways to perform in each, and many alternative paths to success in the overall enterprise. Although in general successful discoverers perform some activities and heuristics more often than those who are unsuccessful or inefficient, there appears to be no pattern of strategies that guarantees success (e.g., Schauble et al, 1990; Shute et al., 1989), a fact that undoubtedly contributes to the lack of consistent patterns of student activity found in this study.

Other researchers have sugggested that the development of scientific reasoning entails not only mastery over particular inference strategies, but also increasing ability to coordinate one's existing theories with patterns of evidence (Klahr & Dunbar, 1988; Kuhn, 1989). Improvement in this coordination of strategies is accomplished by coming to understand the strengths and weaknesses of one's own strategies, and to recognize the occasions and situations when it is appropriate to apply them. Our results suggest that the ability to effect the appropriate deployment and integration of strategies can be learned. With practice, our undergraduates improved in their ability to learn content knowledge from self-directed exploration. That is, they learned more as they proceeded from domain to domain, indicating that they were somehow becoming more effective in planning and carrying out experiments, and in formulating and testing hypotheses based on those experiments. However, as our work probed into the differential complexity and variance of actual domains of science, we have become increasingly aware of the content and context specificity of effective performance. The activity of scientific discovery depends upon variability in the structural form of the discovery task and the nature of the domain.

We found differences in student activity as a function of the particular task and domain characteristics of each of the three computer laboratories. In a situation that required the discovery of correlational regularities, evidence-generation activities, like the heuristic of controlling variables, were primary. Where subjects held prior misconceptions, controlled experiments were essential if biases were to be overcome. In discovery situations where the regularities were functional rules, evidence interpretation became important. When the rules were quantitative, mathematical and algebraic heuristics were particularly strong abilities.

What do these findings imply for understanding scientific discovery? For a psychology that studies the reasoning of professional scientists, it implies that since most scientists work primarily within the boundaries of their chosen fields and even specialize additionally within those fields, the expertise that develops may be chiefly domain-specific. Sociologists (Latour & Woolgar, 1979) and contemporary philosophers of science (Giere, 1979) confirm that for professionals or journeymen practitioners, acquiring good scientific skills is chiefly a matter of being socialized into the skills of a particular discipline and practice.

The implications are somewhat different if we consider scientific reasoning as part of general education. Given that most students do not become members of any practicing scientific community, much less the infeasibility of introducing them deeply into the practices and reasoning styles common to different topics and domains of science, what do we want students to understand about scientific discovery? It appears that the most generally useful message is that discovery is not a monolithic enterprise where one applies cookbook heuristics described in the standard "scientific method" chapter that begins most secondary school texts. Actual problem solving in science requires adaptibility of reasoning to dom..., roperties. Our point here is not to emphasize extreme domain specificity, nor to discount generally useful strategies. When various discovery contexts are compared, they do have specificities and commonalities. However, when a novel problem is encountered, it is necessary to consider what kind of problem it is, and to apply evaluative and self-regulatory skills to decide which processes and strategies are appropriate to the particular task at hand. That is, the most generally useful heuristics may be those involved in learning to evaluate the discovery and inference requirements of a particular scientific setting.

Our emphasis here is reminiscent of a story told by Schoenfeld (1985) about attempts to teach his university students mathematics not as rote or mindless application of learned algorithms, but as a problem solving activity. Although students had considerable knowledge that was relevant to the solution of the novel problems that Schoenfeld posed them, they did not appear to know when their mathematical knowledge was useful to them, and therefore, they did not always apply the skills they had. Schoenfeld emphasizes the need to be explicit about the applicability of the problem solving heuristics that we teach, including how to evaluate the problem, when to apply a particular heuristic, and how to consider whether alternative appropriate strategies might be available that have their own limitations and advantage for that situation.

The implication is that acquiring reasoning skills *per se* is not sufficient. In scientific reasoning it is important to master the skills in evidence generation and evidence interpretation. Individuals can be skilled or unskilled in this regard, and particular skills are associated with learning success or learning failure in particular contexts. However, as students work to acquire skills in the control of variables, measurement, equation-finding, relations between quantitative and qualitative reasoning, identifying correlations, and the like, they must also learn to evaluate their applicability. For science instruction, the implication is the value of repeated opportunities for self regulation--for practice in specific and varying situational contexts where skills in scientific reasoning can be selectively and adaptively used to discover the kinds of lawful regularities relevant to the principles of a particular domain of investigation.

Footnote

[1]One limitation of this study is that the analysis we have completed so far focuses at the group level. It is likely that strategic consistencies will show up most clearly when we analyze patterns of behavior at the level of individual subjects, an analysis we are now proceeding with.

References

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-125.

Giere, R. (1979). *Understanding scientific reasoning*. NY: Holt, Rinehart, & Winston.

Klahr, D., & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science 12*(1), 1-48.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4),674-689.

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science 12*, 139-175.

Latour, B. & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Beverly Hills, CA: Sage Publications.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery: Computational exploration of the creative processes*. Cambridge, MA: The MIT Press.

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1990). Causal models and experimentation strategies in scientific reasoning. Pittsburgh, PA: LRDC Technical Report.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, Fl: Academic Press.

Shute, V., Glaser, R., & Raghavan, K. (1989). Inference and discovery in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.). *Learning and individual differences: Advances in theory and research* (pp 279-326). NY: W. H. Freeman.

Table 1

Student Activities that Differed Significantly Across Laboratories*

Activity	Voltaville	Refract	Smithtown
GENERATION OF EVIDENCE			
Amount and breadth of search:			
Mean number of experiments run	15.0	31.0	46.0
Mean number of changes to parameters	6.3	9.1	13.9
Percentage of parameters changed	78.0%	69.0%	38.0%
Mean number of changes made to variables	6.9	22.7	25.2
Informativeness of search:			
Percentage of minimal required evidence	95.8%	80.8%	50.5%
Structure of search:			
Percentage of controlled experiments	30.4%	82.5%	88.5%
INTERPRETATION OF EVIDENCE			
Making predictions:			
Percentage of experiments with predictions	74.9%	73.5%	53.4%
Effects of prior knowledge:			
Number of alternative hypotheses stated	11.7	12.1	17.2
Percentage of goals discovered	88.5%%	58.3%	52.1%
DATA MANAGEMENT			
Percentage of experiments recorded in notebook	91.4%	99.3%	92.1%
Number of tables created	0.9	3.3	2.3

Note. ANOVAs performed on each of these measures are significant, $p < .05$.