The Implications of Project-Based Pedagogy for the Classroom Assessment Infrastructures of Science Teachers

Laura D'Amico[1]
Learning Research and Development Center
University of Pittsburgh

1999 Annual Meeting of the Educational Research Association
April 19-23, 1999
Montréal, Québec, Canada

ABSTRACT

In project-based science teaching, teachers engage students in the practice of conducting meaningful investigations and explanations of natural phenomena, often in collaboration with fellow students or adults. Reformers suggest that this approach can provide students with more profitable learning experiences; but for many teachers, a shift to such instruction can be difficult to manage. As some reform-minded teachers have discovered, classroom assessment can serve as a vital tool for meeting the challenges associated with project science activity. In essence, it functions as an infrastructure that both students and teachers rely upon as a mediational tool for classroom activity and communications. The research described here investigated the classroom assessment infrastructures created by three teachers involved in the Learning through Collaborative Visualization (CoVis) Project from 1993-94 to 1995-96. Each of the three teachers under study either created a new course or radically reformulated an old one in an effort to incorporate project-based science pedagogy and supporting technologies. Their instructional goals and the role projects played within them varied widely, as did their assessment infrastructures. This paper provides a brief case study of each course which outlines the complex relationship between their curricular and assessment designs. In addition, a cross-case analysis demonstrates the common challenges faced and solutions employed by the teachers as they crafted assessment infrastructures in support of project-based science activity.

[1]The research reported here was originally completed as part of my dissertation work as a graduate student involved in the Learning through Collaborative Visualization Project at Northwestern University. That project was funded by NSF grants RED-9454729 and MDR88-55582 as well as the Illinois State Board of Education under the Eisenhower program. I'm deeply grateful to the teachers and students who participated in the study and the members of my dissertation committee who supported the work: Louis M. Gomez, Roy D. Pea, Allan Collins and Brian Reiser. I also extend my gratitude to Joe Polman, Kevin O'Neill and Nancy Israel for their useful and insightful comments on early drafts of this document.

INTRODUCTION

In the last decade, traditional forms of science education have come under increasing criticism. Researchers and practitioners alike note that a reliance on textbooks, lectures and highly structured labs as vehicles for teaching science can lead to the development of “inert” forms of knowledge (Whitehead, 1929) which students are unable to apply in appropriate contexts (Brown, Collins & Duguid, 1989; Hawkins & Pea, 1987; Lave, 1990), as well as a misunderstanding of the socially constructed, complex and dynamic nature of scientific knowledge and practice (Bazerman, 1988; Carey, 1989; Latour & Woolgar, 1979; Lederman, 1992; Linn, 1991; Myers, 1990). As the American Association for the Advancement of Science warns, “Students cannot learn to think critically, analyze information, communicate scientific ideas, make logical arguments, work as part of a team, and acquire other desirable skills unless they are permitted and encouraged to do those things over and over in many contexts,” (AAAS, 1990, p. 199). These concerns have led the development of new standards and guidelines for science education (e.g. AAAS's Benchmarks for Science Literacy, 1990; Science for all Americans, 1993; and NRC's National Science Education Standards, 1996) which emphasize the need for inquiry-based learning environments where students can build their scientific understandings as they work with others to make meaningful investigations and explanations of natural phenomena. Classroom instruction designed along these lines often places teachers in the role of facilitators who help students build from their current knowledge and proficiencies toward the development of key disciplinary ideas and practices in science.

For many teachers, a shift to instruction of this kind constitutes a radical departure from the customary teaching and learning activity of their classrooms. They face enormous challenges as they not only reconstitute their own practices and expectations, but also those of their students. As some reform minded teachers have discovered, classroom assessment can serve as one of their most vital tools for meeting these challenges. Teachers' assessment practices not only allow them to evaluate what students know and can do, but also serve as a crucial tool for structuring classroom activity and communicating expectations to students. By specifying both the shape and the “worth” of specific intermediate and final products, as well as the behavior or effort they expect to lead to those products, teachers indicate the kinds of work and activity which are valued in class. In essence, assessment forms an infrastructure that both students and teachers rely upon as a mediational tool (Wertsch, 1991) for classroom activity and communications. Teachers construct such infrastructures with the artifacts and behaviors they require of their students; the criteria they use to judge them; the relative value they place on each artifact or behavior; and the rules and customs they and students follow for exchanging assessment related artifacts and information. Since both classroom activity and teacher expectations change as a result of a switch to the kinds pedagogical practices described above, teachers enacting these reforms need to make concurrent shifts in their classroom assessment infrastructures which will enable them to guide as well as evaluate new kinds of student work.

The Learning through Collaborative Visualization Project

The study reported here investigated the classroom assessment infrastructures created by three experienced, reform-minded, high school science teachers to meet the challenges of the current suggested reforms in science education. All three were involved in the NSF funded Learning through Collaborative Visualization (CoVis) Project, a six year collaboration involving nearly 150 teachers nation-wide and research staff at Northwestern University in an effort to improve earth and environmental science education through the use of project-based science pedagogy and supporting technologies (Pea & Gomez, 1992; Pea, 1993; Pea, Gomez & Edelson, 1995; Fishman & D'Amico, 1994; Gordin, Polman, & Pea, 1994; Edelson, in press; O'Neill & Gomez, 1994). Begun in 1992, the six year project had two stages. The first, a "proof of concept" phase, involved six teachers in two Chicago area suburban high schools. During the second scale up phase, the project grew to include nearly 150 teachers in over 40 high schools and middle schools across the United States.

Communities of practice (Lave & Wenger, 1991) served as the conceptual model underlying the form of project-based science pedagogy employed by teachers involved in the CoVis project. In essence, communities of practice are groups of people who share similar goals and interests. In pursuit of these goals and interests, they employ common practices, work with the same tools and express themselves in a common language. Through such common activity they come to hold similar beliefs and value systems.

In the CoVis Project, the process of designing and carrying out science projects was used as a vehicle to engage students in the central activities of the community of scientists, such as asking questions about the world in which they live, designing investigations, working with others to seek answers, and crafting conclusions about the insights they had gained. Computing and telecommunications technology was used to provide students with access to scientific data and analysis tools, as well as to allow them to communicate with one another and other members of the professional scientific community. In combination these technological and pedagogical changes enabled the students involved in the project to come in contact with the values, goals, language, tools, practices and people of science.

Assessment and reform

The observation that assessment and reform are tightly coupled is not new (Kulm & Malcom, 1991), however most of the scholarship exploring this link has concentrated on the role of large scale assessment. Citing the powerful guiding role assessment plays within the educational system, these reformers argue that curricular and pedagogical change is only possible with concurrent changes in educational assessment (Fredericksen & Collins, 1989; Wiggins, 1993). Their arguments are based on the historical effects of large-scale, high-stakes standardized tests (e.g. SAT's & Iowa Basics) on classroom instruction and student learning in which students, teachers and schools adapt their teaching and learning strategies to maximize test scores (Resnick & Resnick, 1991). Proponents of assessment reform seek to mold the existing testing system to their purposes by replacing these content of the tests (e.g. Frederiksen & Collins, 1989; Guskey, 1994; Shavelson, Baxter & Pine, 1992; Wiggins, 1993). Teachers will still "teach to the test", but it will be a new test, one that engenders the desired educational reforms (Darling-Hammond, Ancess & Falk, 1995; O'Neil, 1993; Vickers, 1996). Measurement specialists have worked diligently at creating such new tests (e.g. Lane, Parke & Moskal, 1992; Messick, 1994; New Standards Project, 1997; Shavelson & Ruiz-Primo, 1998) and researchers have begun to evaluate their implications for classroom practice (e.g. Stecher, et. al, 1998; Borko & Elliott, 1998).

The body of work on the relationship between classroom assessment and reform is a smaller one. Historically, the educational assessment literature has contained an overall paucity of research into classroom assessment (Stiggins & Conklin, 1992). However, work of the past decade or so has included research on classroom assessment practices in general (e.g. Airasian & Jones, 1993; Briscoe, 1994; Borko, Flory & Cumbo, 1993; Brookhart, 1993, 1994; Frary, Cross, & Weber, 1992; Shepard, 1997; Stiggins, Frisbie, & Griswold, 1989) and the role or influence of school or district grading policies on them (e.g. Austin & McCann, 1992; Cizek, Rachor & Fitzgerald, 1995); the development of alternative classroom assessment techniques to better support teaching and learning (e.g. Darling-Hammond, Ancess, & Falk, 1995; Herman, Aschbacher, & Winters, 1992; Marzano, Pickering, & McTighe, 1993; Wolf, 1989); studies of assessment responses to educational reforms (e.g. Saxe, et. al, 1997); and investigations into the relationship between classroom and large-scale assessment reforms (e.g. Freedman, 1993; Sheingold, 1995). The study described here places unique emphasis on the use of classroom assessment as a design tool for supporting pedagogical reforms, rather than as a reform in and of itself, or as a reaction to the larger assessment reform movement.

Assessment as an infrastructure

Viewing assessment as “the underlying foundation or basic framework (as of a system or organization)” (Webster's 9th Collegiate Dictionary, 1991) of the educational system allows us to think of classroom assessment as a local piece of a wider infrastructure, that is both supported and constrained by that wider infrastructure. Other researchers have used different terms and metaphors to describe the dynamic and pervasive role assessment plays in educational systems. Examples include Frederiksen and Collins' systemic view of educational testing (1989) or Stiggins and his colleagues description of classroom assessment as an environment or ecology (Stiggins & Bridgeford, 1985; Stiggins & Conklin, 1992). Thinking of assessment as an infrastructure does not contradict either of these other useful images and frameworks. However, it is particularly evocative of some of the traits and purposes of assessment that are salient to teachers intending to reshape the curricular and pedagogical form of their classrooms and to researchers trying to understand the design decisions teachers make. Infrastructures have three characteristics that are particularly germane to the consideration of assessment. First they support and constrain activity in socially acceptable ways. Second, they are often noticed only when they are no longer functioning adequately. Finally, while extending or circumventing infrastructures is usually fairly easy, restructuring is often difficult, expensive, time-consuming, and sometimes requires cultural shifts that are not easy to make.

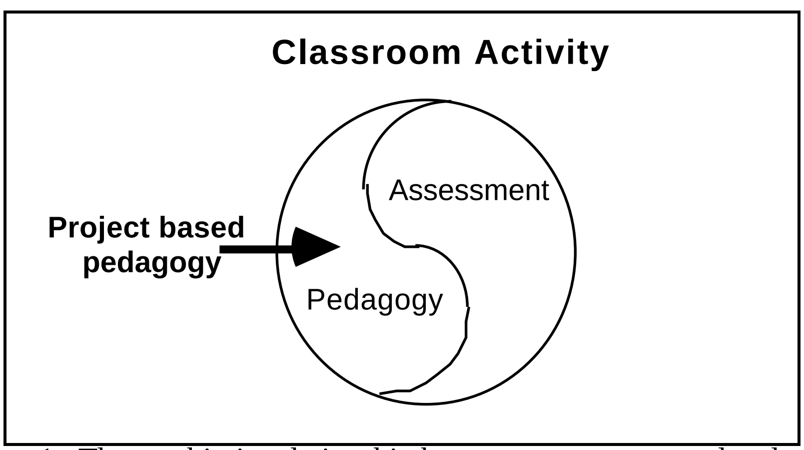

Figure 1: The symbiotic relationship between assessment and pedagogy

Figure 1 above represents how pedagogy and assessment within a classroom are mutually supportive activities. Like other systems, classrooms tend toward an equilibrium. Thus, changes to one component (e.g. pedagogy) must be matched by supporting changes in the other components (e.g. assessment) in order to create a new equilibrium, or else the system itself will wash away those changes as it gravitates towards its original equilibrium (Frederiksen & Collins, 1989). Project pedagogy adds a potentially reforming influence into the classroom system. In order to create a new equilibrium, the classroom assessment infrastructure needs to be changed so that it is supportive of the new activity. The wider assessment infrastructure of the school or community (e.g. grading and tracking systems or state mandated tests) provides a set of constraints that limits teachers' ability to radically reconstruct they way they do assessment. For example, if the school requires a final exam, then they must give one. Such assessment constraints, in turn, limit the teachers' ability to make pedagogical changes. For example, if all classroom learning is structured around students doing individual projects on topics of their own choosing, from what common knowledge does a teacher create exam questions?

This paper presents findings from a study of the complex relationship between pedagogical reform and classroom assessment design as seen through case studies of the endeavors and reflections of three experienced CoVis teachers (D'Amico, 1999). A cross-case analysis of the project-based curricula and assessment infrastructures of these three teachers showed that four aspects of project-based pedagogy dominated their assessment design decisions: (1) the long-term, open-ended nature of project work; (2) the cooperative and/or collaborative nature of project work; (3) the resource intensive nature of project work; and, (4) the iterative nature of project work. The assessment infrastructures built by the teachers to meet the evaluation and support challenges created by these aspects of project pedagogy differed, despite some similarities in the basic building blocks used. In addition to reviewing the synergistic relationship between the design of pedagogical reforms and classroom assessment infrastructures, this paper discusses the effectiveness of those designs in the eyes of the teachers and their students.

METHODS

Data in the form of interviews, observations, surveys, student work and teacher records were collected from the three teachers and their students over a two and a half year period (January of the 1993-94 school year until May of the 1995-96 school year). All data collection and analysis was conducted by a single researcher conducting deliberate inquiry (Erickson, 1986). From these data, a detailed interpretive case study was produced for each of the assessment infrastructures created by the three teachers. Afterward, a set of cross-case analyses was written using common themes that emerged from all three cases (D'Amico, 1999). The findings reported here are based largely on the teacher and student interview data, along with the teachers' grading records. A small number of survey questions are used here to contextualize findings from the interviews, and brief vignettes from classroom observations are sometimes used to illustrate them. An abbreviated form of each case study is provided here, as well as a summary of the subset of cross-case findings that directly relate to the relationship between project-based science and classroom assessment design.

Data collection and analysis

Teacher Interviews

A total of 33 interviews, eleven per teacher, were conducted with the teachers over the two and half year period. Ten of these interviews used an “interview guide approach” (Patton, 1990) in which general questions were laid out in an interview guide, but the exact order and wording of those questions were decided upon during the course of the interview and interesting emergent themes were followed at the interviewer’s discretion. These ten interviews fell into two categories. The first two interviews in the winter and spring of the 1993-94 school year were exploratory in nature and quite unstructured. The first interview focused on understanding each teacher’s concept of the characteristics of a good science project. The second interview explored each teacher’s beliefs about assessment and project-based pedagogy more broadly, touching upon both formal and informal assessment practices. The other eight interviews were more structured and occurred after every marking period during the 1994-95 and 1995-96 school year. They focused on developing a detailed depiction of the grading practices used during that academic quarter, as well as a review of classroom activity and goals for the same period. Gradebook records were collected and reviewed at the time of the interview. The eleventh interview was a “standardized open-ended interview” (Patton, 1990) on the instruction goals of each teacher conducted in the early fall of 1995-96. From previous interviews, a list was generated of the types of instructional goals the teachers typically mentioned. The teachers were asked to rate on a scale of one to ten how important goals in each category were to them and how effective projects were at helping them meet goals of that type. They were also asked to explain their reasoning behind each rating and to provide examples of their personal goals which fit into each category.

Student Interviews

Twenty-eight student interviews were also conducted during the winter and spring of 1995-96. Most of these were with a small subset of students chosen for closer study. (See the section on

“Sampling” below.) An attempt was made to interview each of these “focus students” three times—once in the winter (January or February), once after their third quarter project was complete (April), and once just before school let out (May or June). Two of these interviews—the winter interview and the spring interview—covered essentially the same topics. They asked students to explain the purpose of the course, what they thought their teacher wanted them to learn, what it took to do well in the class, how grades were calculated, and what kind of feedback they got on their work. The project interview asked each student to review the experience of doing their third quarter project: their understanding of the assignment, the kind of work they did to complete it, how much time they put into the work, and how good they thought the final results were from both their own and their teachers' perspectives. A total of thirteen students were each interviewed between one and three times. All three of the interviews used a “standardized open-ended interview” format (Patton, 1990).

Data Coding

All interviews were audio taped in addition to being recorded through field notes. All interviews with both teachers and students from 1995-96 along with teacher interviews conducted in the winter of 1993-94 and fourth quarter interviews of the 1994-95 school year were transcribed. All interview notes and transcript data were imported into NUD*IST™ 4.0 (QSR, 1997), a qualitative data analysis software package, and coded in three ways. First, all the “standardized open-ended interviews” were indexed by question so that the responses of interviewees to the same question could easily be retrieved and compared. Second, the interviews of each teacher and their students were coded according to themes and issues that were particular to each course and had emerged through early work on the case studies. This enabled the comments of the teacher and his/her students to be compared and contrasted within each case. Finally, all interviews were coded on clusters of cross-case themes which had emerged through the course of the study so that they might also be retrieved for comparison. The themes included:

(1) four aspects of project-based science pedagogy (long-term/open-ended, resource intensive, iterative, collaborative/cooperative) and their influence on assessment infrastructure design;

(2) the role of technology in each course and its relationship to the assessment infrastructure;

(3) descriptions of the wider assessment infrastructure and its influence on the teachers' classroom assessment designs;

(4) common assessment infrastructure components (e.g. final products, such as papers and presentations; intermediate products, such as milestones and drafts; records and grades of students' work behaviors and attitudes; and conferring systems).

The first and last of these clusters, the four aspects of project pedagogy and common assessment infrastructure components, are the object of this paper's discussion.

Survey data

Survey data was collected from the students in every section of the CoVis-related courses each teacher taught. Three surveys were administered during the year: 131 students responded to the CoVis Fall 1995 Survey; 113 responded to the CoVis Spring 1996 Survey; 102 responded to a “Student Beliefs about Class Goals” survey; and 92 students responded to all three. (See table 2 below.) Descriptive statistics on a subset of the survey items are used here to provide background information on student demographics and attitudes towards science, as well as to demonstrate students’ opinions on the course in question and its assessment infrastructure. The results of these survey items are included in the case study and cross case analyses where appropriate.

Observations

Informal observations as a participant observer and technical assistant began in 1993-94 in all three classrooms. Formal observations (n=102) in which comprehensive notes were kept were not begun until 1995-96. The bulk of these observations (n=72) occurred during the approximately two month period that constituted the third academic quarter in each school. One section of each teacher’s course was chosen for observation. It was presumed that instruction in all sections would be similar, so sections were chosen for the logistical convenience of the observer. (Teacher interview data confirmed the assumed similarity.) For the purposes of this paper, these observations should be considered largely as contextual filters that inform and constrain the analysis of the interview data, rather than as objects of analysis themselves.

Sampling

Teachers

The three teachers[2] chosen for this study had been with the CoVis project since its inception in 1992. They each worked at one of two high schools in the suburban Chicago area—Roger Wolfe and Carol Patterson at Edgewater High School and Gary Magi at Lakeside High School.[3] Roger and Carol shared the "CoVis classroom" at Edgewater with a third teacher. This room had six computer stations in it, all of which were connected to the internet and two of which were attached to a desktop teleconferencing system. The room also had its own printer. Gary initially shared the "CoVis classroom" at Lakeside with two earth science teachers, but in 1995-96 moved to a new, much larger room that could handle the increased enrollment in his class, but did not have computers in it. Instead, he regularly used the school's central computing center, which had one lab of computers connected to the internet, for access to computing and telecommunications technology.[4]

All three teachers went through a period of intensive course design between the 1992-93 and 1994-95 school years. Gary and Carol began building brand new courses in 1992-93: science, technology and society (STS) and Environmental Science, respectively. Roger, on the other hand, started with the same class he had taught for over 20 years: earth science. In 1992-93 he began to redesign both the purpose and structure of the course from scratch. All three are competent teachers and so by 1995-96, the focal year of this study, they had each built stable project-based curricula and associated classroom assessment infrastructures for their courses.

The teachers' evolutionary experiences in designing and redesigning these science courses made them ideal cases to study. The immense and immediate challenges they faced during their restructuring efforts brought issues of assessment strongly to the foreground. Further, they were

[2]The data collection initially involved all six original CoVis teachers. Midway through the 1994-95 school year, this number was reduced to four in order to make data collection more feasible. The four teachers included two earth science teachers, one environmental science, and one science, technology and society (STS) teacher. Due to time constraints, the data analysis and conclusions drawn here are based on only three of these final four teachers. The most traditional of the four, one of the earth science teachers, was removed from the final analyses. However, since data from her class was included throughout the data collection and initial analysis stages, her work and that of her students were an integral part of the thought that went into the design of the analysis structures eventually used with the three remaining teachers.

[3]All names of schools, teachers and students are pseudonyms.

[4]While this paper does not focus directly on the role technology played within the teachers' pedagogical and assessment designs, it is difficult to understand some of their design decisions unless one realizes that all the teachers and students in these courses did have regular access to both computers and telecommunications technology as well as private space on file servers where they could store their personal email communications and ongoing project work.

all reflective and articulate people, making them exemplary informants on the change processes they were experiencing. Finally, despite their common experiences with both CoVis and the challenges of creating new courses, they ranged widely in their teaching experience, style of project based science implementation, grading practices and instructional objectives. This diversity allowed for an exploration of the different forms that assessment infrastructures take when mature science teachers enact project-based pedagogy under a range of conditions and for a variety of purposes.

Teacher	Years teaching	CoVis classes	Periods per week	Other classes or duties	Total classes
Gary Magi	14	Science, Technology and Society (STS) (2 sections)	2 periods/day for each section=20	Conceptual Physics (1 section)	3
Carol Patterson	5	Environmental Science (1 section)	7 periods per week	Chemistry (3 sections)	4
Roger Wolfe	23	Earth Science (3 sections)	7 periods per week for each section=21	CoVis technical coordinator	3

Table 1: Teaching experience and loads as of the 1995-96 school year

The three teachers had a range of teaching experience. (See table 1 above.) In 1995-96, Roger had been teaching at Edgewater since he left college twenty-three years earlier. While he had occasionally taught chemistry, earth science was his passion and eventually led him to obtain a masters degree in geology. At the other extreme, Carol had just been teaching for five years in 1995-96, during four of which she had been involved as a member of the CoVis Project. An environmental engineer and chemist by training, she spent five years in industry before beginning a new career in teaching. Gary, who has an undergraduate degree in physics as well as a masters in mechanical engineering, worked as a mechanical engineer specializing in energy conservation and passive solar design for eight years before becoming a teacher. An experienced teacher of fourteen years in 1995-96, he had taught in several places, including Lakeside for the last nine years, a German gymnasium for two years, and in the Chicago Public Schools for three years. Throughout his teaching career, he taught primarily physics, though in Germany he taught math as well. While with the Chicago Public Schools he taught chemistry on occasion. All three of these teachers used their significant experience with research in science as well as applications of science to inform their instructional designs.

Students

There were a total of 203 students enrolled in the CoVis-related courses taught by the three study teachers in 1995-96. (See table 2 below.) Roger Wolfe had 85 students in three sections of earth science during the 1995-96 school year. Carol Patterson had 20 students enrolled in her single section of environmental science. Finally, Gary Magi had 98 students enrolled in two, double period sections of science, technology and society (STS). Both teachers at Edgewater had fairly ethnically homogenous classrooms (78% white, 1% black, 21% other, n=76), which reflected the composition of the school as a whole (88% white) and its catchment community (93% white).[5] Gary Magi's students (51% white, 24.5% black, 24.5% other, n=45) reflected the somewhat wider diversity of the Lakeside's school (50% white, 42% black) and community (70% white, 21% black) populations. In general, the students in all three classes came from families that valued education. The mothers of the majority of the students had at least a bachelor's degree (60% of Roger's students, 80% of Carol's students, 64% of Gary's students) and many had graduate degrees (21% of Roger's students, 33% of Carol's students, 45% of Gary's students). A similar pattern existed for their fathers. Very few of the parents (two mothers of Roger's students and three fathers of Gary's students) had not completed high school.

Teacher and class	**Fall Rosters**	**Fall Survey**	**Spring Survey**	**Goals Survey**	**All Three**
Roger Wolfe–Earth Science	85	67	53	44	42
Patterson—Environmental Science	20	15	15	15	14
Gary Magi–STS	98	49	45	43	36
Total	**203**	**131**	**113**	**102**	**92**

Table 2: Number of students in each class

With the advice and cooperation of the teachers, a small subset of these students were chosen for close observation and study. Teachers were asked to identify from the section of the course being observed a strong student who was doing particularly well in the class, a student whose work was representative of the "average" student in the class, and a student who was struggling. These nine "focus students" were each asked individually if they were willing to be studied more closely and informed that this closer study would include one-on-one interviews, copying some of their class work, and closer observation during class. All of the original twelve asked agreed, though one was eventually dropped from the study after being suspended and permanently removed from Gary's class. In each class, one to three extra students were selected with assistance from the teacher and interviewed at least once (in one case twice) in order to create a broader set of perspectives than

[5]Demographic data for students in each course were obtained from their responses on the CoVis Fall survey. Demographic data for the entire student population at Edgewater and Lakeside, along with those of the communities in which each school resided, were obtained from the Sunspace Internet site (www.sunspace.com) in the fall of 1996. The data sources used by this site included: U.S. Census Bureau, 1990 School District Special Tabulation; National Center for Educational Statistics, U.S. Department of Education; Administrative data is from individual school systems. All data are from the 1990 census or pertain to the 1989-90 school year.

those held by the original focus students. In all cases, these extra students had strong feelings about the course, either negative or positive, that differed from those of the other students being interviewed.

RESULTS

In 1995-96, each of the three study teachers—Carol Patterson, Gary Magi, and Roger Wolfe—were in their fourth year of involvement with the CoVis Project and at least their third year of project-based science teaching. All three believed in the power of projects to create powerful learning experiences for their students. However, since their instructional goals differed, the kinds of projects their students did and what they were meant to learn from them differed as well. Some of these differences were based on the differences in the subject matter they taught and the core purpose of their course. Others were based on their beliefs about teaching and learning.

The classroom assessment infrastructures created by all three teachers to meet the challenges of project-based science were based upon variations of four basic design components: final products or performances, intermediate products, records of student daily work and activity, and face-to-face meetings or conferences with students. Despite these similarities, the particular shape each of these components took in the four classrooms and the ways in which they were combined to create each assessment infrastructure differed. Dimensions of variation included the amount of systematicity used to collect data through any of these three sources, as well as the regularity of their use in providing feedback to students and the centrality of their importance in constructing students' grades. Below is a brief description of each teachers' assessment infrastructure followed by a cross case comparison of the way these infrastructures interacted with the key aspects of project-based science pedagogy.

Earth Science at Edgewater

There were two goals or principles which underlay the entire design of Roger Wolfe's course, including its assessment infrastructure. First and foremost his course was designed to teach students about science by engaging them in the practices of science. Second, his course was designed to maximize student learning by providing students with the opportunity to study areas of interest to them. His goal was for his students to learn deeply about the concepts and issues in a few areas of earth science of interest to them and to understand and be able to practice key activities in the field of science, such as analyzing data in order to draw and report conclusions. These learning goals and pedagogical principles led to several instructional needs. First, his students needed to learn enough about the field of earth science to both develop interest in some of its issues and topics and craft questions worthy of investigation. Second, they needed to learn enough about the process of science to recognize the feasibility of investigating their question given the data and analysis means at their disposal, as well as design and conduct that investigation. Third, they needed to learn to use the scientific tools they had at their disposal. Finally, they needed to learn enough about the language and forms of communications in science that they could produce reasonable written and verbal descriptions of their work. To meet these needs, Roger structured his course so that students received one academic quarter of introductory material and three quarters of open-ended project work.

During the introductory quarter of 1995-96 students received a whirl-wind tour of the major topics and concepts in Earth Science through a series of lectures, videos and demonstrations, while at the same time they learned to use the computing equipment and software available in the room. During the project quarters, students collaborated in groups of two or three on long-term scientific investigations of their own choosing in areas related to earth science. While the students had quite a bit of freedom with respect to the topic or phenomena they chose to study, Roger was quite specific about other aspects of their work. In particular, he did not want students to write "informational" papers that simply summarized what was already known about a topic from books or periodicals. In order for a project in Roger's class to be a *science* project, "You have to have a scientific question to answer, and then, you have to collect data to answer that question. You have to analyze your data to see what it means, and then you come to a conclusion" (R. Wolfe, interview, March 10, 1994). All the projects in Roger's class culminated in a final paper and accompanying class presentation.

Roger created two key documents that outlined his expectations for students' work in the course and provided a set of directions for doing projects: a detailed two-page description of the format students were supposed to use for their final paper entitled "Project Reports" and a set of instructions entitled "How to do an Earth Science Project". Since students could study any topic they wished, the guidelines and instructions Roger provided them for doing their work were necessarily somewhat generic. As a result, a significant portion of his assessment infrastructure was designed to provide opportunities for him to gage the progress of individual groups and provide them feedback specific to their project topic.

The components of Roger's assessment infrastructure during the first quarter of the year, when students were engaged in learning introductory materials, included a test on the basic concepts and processes in earth science covered in his lectures, a computer skills performance test, two technology assignments and a work grade. The tests and assignments he gave were meant to both encourage students to take the introductory activities seriously as well as ascertain which students might need further information, practice or support. The work grade was meant to encourage students to be productive during class time set aside for learning to use technology.

The key components of his assessment infrastructure during the other three quarters of the year were a final paper and presentation, a work grade, and a set of project milestones for each of the three project cycles. The purpose of the final paper, worth between 25% and 50% of the students' grades depending on the project cycle, was to demonstrate what the students had learned in the course of their investigation and to give them practice writing in the genre of the scientific research article (Swales, 1990). The presentations, worth between 15% and 25% of the students' grades for each project cycle, were meant to expose students to a wider range of topics in earth science (since each of them would only study three in depth) and to develop their ability to critique other students' work. The work grade portion of Roger's assessment infrastructure, worth between 19% and 33% of the students' grades each quarter, was based upon his daily records of students use of class time and students' ratings of their own and their teammates' contributions to the project. It was meant to both encourage students to use their class time well and to give credit to students who did more than their fair share of their project team's work.

The milestone portion of Roger's assessment infrastructure, worth between 17% and 19% of each student's grade, served several purposes. First, as shown in table 3, the project milestones helped students break down the activity of doing a science project into logical subtasks. Second, some of those subtasks (e.g., background information and data/analysis) resulted in written text that could be incorporated directly into the final paper (e.g., introduction and methods/results), which made the logic behind accomplishing them compelling to the students. These were not "busy work" activities—they were helpful toward accomplishing their final goal. Third, the milestones also helped students to pace their efforts, since the intermediate deadlines prevented students from procrastinating on their project work until just before the final deadline. Finally, each deadline served as a pressure point that encouraged students who were having difficulty to seek advice from Roger. In addition, Roger would usually quickly hand back the written pieces, such as the background information, with copious notes on them. Thus, the milestones afforded an occasion for students to receive both verbal and written support and feedback tailored to their specific project topic.

Project subtask	Associated milestone
1. Assemble a project team and collectively agree on an area or issue in earth science that you which you all want to study	Turn in your topic and the names of your team members
2. Lookup background information on your topic so that you can ask a researchable question	Turn in a summary of the background information you read
3. Decide on a research question or proposal	Turn in a question or proposal
4. Collect or find data to answer your question and analyze it	Turn in your data and data analysis (These were sometimes divided into two separate milestones.)
5. Come to a conclusion that answers your question based on findings from your data analysis and write a final report describing that conclusion	Turn in your final paper

Table 3: Project subtasks and associated milestones in Roger's course

Every academic quarter, Roger built a spreadsheet for calculating his students’ grades. His calculation methods were somewhat idiosyncratic and could best be described as a series of embedded relative percentages. Everything he graded initially started out with a point value, which he then converted to a percentage. He then weighted the percentage scores for each graded item either by combining it with other items (e.g., adding students’ work grades to their milestone grade and dividing by two) or by specifically adding a weight to the calculation (e.g., multiplying Roger’s assessment of the students’ work habits by three, adding to it the students’ assessment of their work and dividing by four). Through a series of four or five such combinations Roger finally determined each student’s quarter or semester grade. A copy of the spreadsheet, with each student identified only by their school identification number, was available via the CoVis file server for students to access at any time. It was updated after each milestone or final product was collected and evaluated.

In sum, assessment in Roger’s class, was closely tied to practice. “How do you measure how somebody thinks? How well they’ve developed the capacity to think?” he asked and then added wistfully, “If you put some sort of an electrode cap on their heads to find out what they’re learning, while they’re doing it, that would be a thing to have for assessment,” (R. Wolfe, interview, June 2, 1995). In the absence of such a device he felt his best alternative was to “set up the experiences and then evaluate those” (R. Wolfe, interview, June 2, 1995).

> I have this gut reaction that [...] the best thing I can do is kind of lead them down the right path to produce the kinds of things that, or set up the experiences for them to do activities and give them the opportunity to produce artifacts like this paper that would involve those skills that I'm trying to get them—skills or what ever you want to call them—and then at least grade those things (R. Wolfe, interview, June 2, 1995).

Environmental Science at Edgewater

The goals and principles behind the design of Carol Patterson's course led it to have a somewhat more traditional structure than Roger's. Faced with many students who were both passionate and somewhat naive about environmental issues, Carol's primary goal was for them to become literate and critical consumers of scientific information. In particular, she wanted them have enough knowledge of the *science* behind the arguments of environmental policy makers to make informed decisions as citizens and voters. She wanted them to be able to listen to, reason about, and integrate the evidence provided by all sides of the often heated environmental debates. As she explained in an interview, "I feel like already from the beginning that the course was created for a lot of reasons. Mainly I wanted them to be literate about environmental issues and be able to talk from two sides of an issue and really be rational," (C. Patterson, interview, March 10, 1994). Second, like Roger she believed that students learned best when they could explore issues of interest to them. However, that belief was tempered by both a desire that all students leave the class with a common core of knowledge about certain environmental processes, and a conviction that many of her students needed more instructional support than a pedagogical design like Roger's could provide them. She said that typically the students who took her class had, "no tolerance for ambiguity. They [could] not stand not knowing what to do when" (C. Patterson, interview, September 28, 1995).

These learning goals and pedagogical principles led to several instructional needs, some of which were similar to Roger's and some of which were different. First, her students needed to learn what the key debates in the field were, what positions scientists and policy makers tended to take with respect to those debates, what arguments each side made, and what evidence supported their arguments. Second, they needed an understanding of the environmental mechanisms and processes (e.g. the water cycle and the carbon cycle) associated with those debates. Finally, they needed to understand the research tools and processes used to collect the evidence which supported the arguments on each side, as well as the politics behind funding and conducting environmental research. As she explained in one interview:

> I just think that what scientists do is a black box to students. They think they're just a bunch of white men in white coats and they sit around and...tell the world what's happening. I think they think it's a pristine, objective thing that's going on and...that

> and everyone accepts it, that the money to fund research comes from out of the blue, (C. Patterson, interview, September 28, 1995).

To meet these needs, Carol designed her course as a combination of the activities traditionally found in a science class (such as lectures and labs), and non-traditional ones (such as long-term projects and reaction essays about articles presenting controversial views on environmental issues). Recognizing that many of her students were as interested (or more interested) in the social and political implications of environmental issues as the science behind them, Carol made environmental decision making the focus of the course. The activities of the course were thus all woven around units that focused on key environmental topics or debates that received significant media attention, such as global warming, the potential degradation of the ozone layer, and concerns about the proper disposal and management of various kinds of waste. Topics such as these were all of concern and interest to the students and at the same time central to the field.

While some of the topics, labs, activities, project cycles and articles have changed over time, this basic curricular structure remained essentially the same between 1992-93 and 1995-96. In the 1995-96 school year the major topics of study included water quality, land use management, toxicology and risk assessment, the carbon cycle and global warming, ozone depletion, plastics and persistent and hazardous wastes. Each project lasted between four and eight weeks. Projects were conducted one per quarter in the following order: create a Land Use Management Plan (LUMP) for a known city (structured group project); assess the risk levels in your diet due to food additives (structured independent project); explore a topic in global warming (open group project); explore a topic in environmental science (open independent project).

The project work, which in 1995-96 accounted for 61% of the students' first semester grades and 48% of their second semester grades, had various purposes depending on the project cycle and its place within the course. Carol had students create Land Use Management Plans (LUMPs) early in the year in order to demonstrate to them the complexity of environmental decision making—that it must take into account the economic, social, and political needs of people as well as environmental conservation concerns. The Food Additives project made the processes of, and debates around, risk assessment both concrete and personal. The Global Warming and Independent Projects were meant to give students an opportunity to explore an issue of concern to them in depth.

The first two projects Carol did in 1995-96 were both fairly structured. The students had quite a bit of room to be creative in their LUMPs. However, since the purpose of the project was for students to integrate and juxtapose the multiple demands placed upon land use, Carol provided firm parameters for the issues they must address in their proposals. Likewise, the food additives project had a set of clear procedural steps: keep a food diary for a week, describe your own eating habits, determine what food additives you consume regularly and then research what potential risk factors are involved in each and determine your personal risk levels. The topics to be covered in the final report were also clearly specified. Later in the year, as she felt the students had acquired more of the skills necessary to conduct independent research projects, she provided them the opportunity to do less structured project work where they choose their own topics and designed their own investigations. The Global Warning project needed to be on a topic related to issues in global warming, but both the specific focus and the kind of study students would do was at their discretion. For their spring independent project they could study any topic in environmental science. During these open projects, Carol encouraged students to do work similar to that of the students in Roger's class—scientific investigations in which the collection and analysis of data to reach a conclusion served as the core activity. However, many of her students proposed (and she accepted) projects more reflective of a literature review or policy statement. In these projects, students typically synthesized and integrated what was already known about a particular environmental concern, often concluding by proposing a plan of action. In these cases, Carol still encouraged her students to grapple with data in some fashion to support their arguments, even if they were not doing an empirical investigation per se.

The assessment infrastructure to support these activities included components related to class work and homework, as well as those meant to support and evaluate project work. Evaluations of worksheets or write-ups associated with labs, thought experiments and related activities composed 17% of the students first semester grade and 23% of their second semester grade. Likewise, evaluations of homework in the form of reaction papers to position articles and review questions associated with textbook chapters, composed 6% to 7% of their grades each semester. Along with tests and quizzes (15% of students' first semester grades and 21% of their second semester grades), these items were graded so that students would take their completion seriously, would study the concepts and processes conveyed through the lectures, discussions, readings, labs and activities, and could be evaluated on their command of those materials. When such evaluations showed that understanding among the class as a whole was limited or incomplete, Carol would attempt to fill the deficit by adding further instruction in the form of lectures, discussions, labs or other activities. The purpose of these activities and the assessment components which supported them was to ensure that all the students in the course understood the essential arguments on each side of the environmental debates being studied, the evidence used to support them, and the environmental mechanisms and processes related to them.

Despite differences in the purpose and shape of each project cycle, there was a consistency to the assessment infrastructure which supported them. For both of the open-ended projects, Carol had students submit a proposal early on in the project cycle which she graded and then discussed with each student team. In addition, for all four project cycles, Carol met with each group on what she called a "touch-base" day two or three times during the course of a project in which students gave her a verbal progress report. Sometimes the students also provided Carol with a written progress report that they discussed during the meeting. These progress reports were sometimes assigned a grade and sometimes simply checked off as complete. The purpose of both the project proposals and progress reports was to help students shape their investigations, pace their work, provide feedback on the quality of their work so far, and facilitate Carol's attempts to provide course correction advice when students were having difficulties.

A final written report of some kind, and in some cases additional supporting documentation (e.g. LUMP required a map and Food Additives required an appendix listing all the additives in the student's diet), was required at the end of all four projects and the specifications for each was provided in advance. For the more open-ended projects, these specifications were necessarily more generic. The students had to give a presentation on the conclusions from their project for all but the Food Additives project. These final products were meant to demonstrate what the students had learned through their project work, and in the case of the presentations, share that information with their fellow students.

Finally, Carol included a "participation" grade each semester that incorporated a number of aspects of students' classroom performance into their grade, such as appropriate use of class time or working relations with their project team. Likewise, she included an "individual" or "personal" component to the final project grade for all group projects. This portion allowed students to be acknowledged for their personal role on the project and differentiated between those who contributed more labor, leadership or vision to the endeavor than others. The participation grades and personal grades were based on Carol's observations of student activity in the class. She took free-form notes on a yellow legal pad which she had with her at all times, as she talked to students and observed them during class. She used these notes as a reference both for tracking student progress and reminding herself of who needed extra assistance as well as assigning personal and participation grades.

Carol used a simple point system for combining the various components of her assessment infrastructure. Each graded item was assigned a specific number of points, and the weight of that item within the students' final quarter and semester grades was based on its point value relative to the those of the other graded items. In sum, her assessment infrastructure reflected the range and diversity of the instructional activities in her course. Items traditionally graded in science courses, such as homework, labs and tests were used to evaluate and support the traditional activities in her class, such as lectures, discussions and labs. Other assessment components, such as proposals, progress reports, conferences with students, papers, presentations and records of student activity, were used to support and evaluate the project-based part of her course.

Science, Technology and Society at Lakeside

The learning goals and pedagogical principles that drove the design of Gary Magi's course in science, technology and society (STS) had some overlap with Carol's and some with Roger's. First, as in Carol's class, the topics and ideas his students explored were a combination of those typically covered in high school science classes with those typically covered in social studies classes. His primary goal was to create a learning environment that would encourage his students to think deeply about the ramifications of the technological and scientific advancements of our age upon the way we live our lives and define who we are:

> Ultimately, I think the central question in the course is what-- is there some way that we can define something called appropriate technology? Is there, is there some way to define a sustainable use of technology? Is, is there a way for us to understand-- to have the criteria to evaluate ethically, financially, socially, politically-- evaluate whether technology is good for us or not? (G. Magi, interview, September 27, 1995).

Thus, in Gary's course, technology is not only an enabler for teaching and learning, it is the primary *object* of study. Second, he believed that the nature of schooling as it was largely practiced at schools like Lakeside was coercive in a damaging way. “Most people leave school thinking that learning is a chore. We've done some serious damage if that's what they believe,” he worried (G. Magi, interview, September 27, 1995). Thus, like Roger, Gary believed that students learn best when they have considerable control over what they learn, “My working assumption is that, for a lot of kids, they will learn more if they are choosing what it is they are learning on a day to day basis, and that's the philosophical bed rock [of my course design],” he explained (G. Magi, interview, June 1, 1995). However, like Carol, he also recognized that many of his students were unprepared to handle the level of autonomy he hoped to provide for them in his course. “The kind of open-ended research that we're talking about here, where they have to figure out how to direct themselves, let alone define their goals—what is it they that they want to do?—is an alien experience for a lot of kids and too much freedom, too quickly, is paralyzing,” he explained (G. Magi, interview, June 1, 1995).

Like Carol, Gary felt his students needed to know what the main issues in STS were, as well as understand the scientific concepts and natural processes related to them. However, unlike Carol, he was less interested in their acquisition of such knowledge as an end in and of itself. Instead, he was more concerned that students have the background knowledge that would enable them to reflect upon and discuss the impact of science and technology on human lives. His students also needed to learn research practices. However, unlike Roger, the research his students conducted more closely resembled that of social science rather than natural science. He said, "We're not doing much of what scientists do. If you brought in the world of scientists to include people who study humans and what they do and machines and what they do, then we're doing a lot of it," (G. Magi, interview, September 27, 1995). Students needed to learn how to convey their reflections on the issues they were studying in a variety of forms, ranging from project reports to web pages to videos, using the conventions of each form. To do so meant that they also needed to learn to use the technologies required to create the various products. Finally, Gary felt that the students in his class needed to learn how to "use freedom to learn" (G. Magi, interview, June 1, 1995).

To meet these instructional needs, Gary designed a course in which project work was combined with other instructional activities tied to a series of topical units, as in Carol's course. "What I'm discovering," he said, "is we need, we need a mix of teacher directed and student directed, of the performance and high energy, everybody doing the same thing at the same time [e.g. lecture, demonstration, whole class discussion] and "Okay, what you are *you* doing today?" you know, student directed activities (e.g. projects)," (G. Magi, interview, November 21, 1995). In an effort to create a balance between project activity and other instructional activities in which students could learn the "basics", Gary and his teaching partner, a social studies teacher named Karen Brandon, created a new and innovative format for the course in 1995-96.

In the past, the course had been divided into days when all the students were involved in whole class activities, such as lectures of discussions, *or* they were all working individually or in groups on their project work. With the addition of Karen as a second teacher for the course, the switch to a double period session and the creation a special research pass system for the course, they were able to rework the course so that there were two simultaneous activities occurring in the course nearly all the time. Lectures, labs, discussions, videos and other large group teacher-directed activities were all referred to as "workshops". Each month Gary and Karen would post a list of the "workshops" that would be conducted, along with handouts describing the preparation work (often reading articles and reflecting on them in a journal) required for entrance into each workshop. On any given day, students could choose either to attend a workshop or work on the their current project. Students were required to attend a certain percentage of the workshops, but which ones they chose were up to them. No more than one workshop was offered each day so that one of the two teachers was always available to help students with their project work. Students working on their projects could stay in the two interconnected rooms that served as Gary and Karen's classrooms or use the research passes to go to the library, computer center or off campus if necessary in order to collect data and information they needed for their project work. In this way, Gary was able ensure that students all received background information on each of topics covered in the course, while still giving them considerable freedom about the particular issues they studied.

The projects Gary's students did, most of which were group projects lasting between two and eight weeks, emanated from the topical units. Like Carol, he designed his course so that the level of student control over the content, direction and form of the final product for each project varied throughout the year. The most structured projects were at the beginning of the year and the most open-ended ones were at the end of the year. Projects were ordered in this fashion specifically to give students the opportunity to "learn to use freedom". The projects in 1995-96 included: (1) a report documenting in detail the infrastructures of each student's home (short, very structured, independent project); (2) a video describing the mechanisms and function a major infrastructure within the school or world at large (long, semi-structured, group project); (3) an essay describing how information technologies affect the life of an adult the students shadowed in their work place (short, structured, independent project); (4) a web page summarizing an issue related to energy and the environment (short, semi-structured, group project); (5) a paper, video, or web page on a topic or issue related to a variety of fields in science and technology including medicine, warfare, mass media, nanotechnology, ubiquitous computing, and the Luddite and Neo-Luddite movements (long, open, group/individual project).

Gary's assessment infrastructure reflected his instructional goals and pedagogical principles. Just as he wanted to avoid coercion in teaching, he wanted to avoid the coercive nature of assessment, "Just the act of assessing is coercion. 'Have you done this yet?' That's a coercive statement. And so, if I'm philosophically opposed to coercion, that raises a very tough question about what the hell assessment is for," (G. Magi, interview, March 14, 1994). His answer was to think of assessment as a "dialog between teacher and student," one which was "as informative to the student as it was to the teacher if it is done right," (G. Magi, interview, March 14, 1994). To this end he used a holistic form of grading which is somewhat unusual in high school courses. Rather than assigning points for every piece of work students did and, and calculating a grade by combining those ratings numerically (as Roger and Carol did), Gary established categories of work and performance for which he holistically graded his students. At the same time, students graded themselves and discussed their ratings with Gary to arrive at a final quarter or semester grade. This discussion was the culmination of the "dialog" that Gary was striving for.

The four categories of work that Gary assessed were: individual work, personal outcomes, project binder, and final team project. Each of these was divided into subcategories. Ratings on the subcategories informed the ratings of the category to which they belonged, but they were *not* combined in an arithmetic fashion. Ratings for both the categories and subcategories were initially on a three point scale (Excellent, Satisfactory, Unsatisfactory), but by the end of the 1995-96 school year they were shifted to the standard five point scale used by the school (A, B, C, D, F).

Gary estimated that between one-third and one-half of each student's grade was based on their project work. A central part of his assessment infrastructure and a key component of each student's grade related to project work were their project binders. Each project team maintained a three-ring binder that contained the complete record of work on the project, including research notes, progress reports, task logs, project milestones (such as outlines or story boards), and drafts. The binders served two major purposes in the course. First, they helped the project team organize and coordinate their work. Second, they helped the project team document and communicate its progress to Gary and Karen. In this sense, the content of the binders served many of the same purposes as the milestones in Roger's class.

Binders were typically graded on the quality of research demonstrated in them, the completeness of the reports and task logs, and the organization and the timeliness of the work within them. As in Roger's class, the milestones were meant to help break down the task of doing each project into its logical subcomponents, which differed based on the project cycle. A milestone common to all projects in Gary's class was the specification of a minimum number of research notes from a minimum number of sources by a set date. Since in many of the projects students summarized, synthesized, or integrated knowledge on a given subject from a variety of sources, the research notes requirement was meant to encourage students to review a wide range of sources rather than simply summarizing one or two main ones.

In addition to milestones, Gary used a form of ancillary reports to shape and guide project work. Task logs were meant to help students plan their work overall, while weekly progress reports were meant to help them reflect on what had been accomplished thus far, the contribution of each project team member to the work, and plans for the next week. Both task logs and weekly progress parts were part of the binders. Gary and Karen held periodic conversations with each team called "binder checks" in which content of the binders, including milestones and ancillary reports served as a reference point for discussing project progress.

The next most significant portion of the students' grades linked to projects was the final product each team completed. It was usually rated on a combination of its technical quality and the depth of knowledge of the subject matter it conveyed. The technical rating encouraged students to take the conventions for that form of communication seriously. The depth of knowledge rating made it clear that their research must have substance to it.

Gary assigned an individual work grade to students by combining individual project work ratings with ratings of workshop performance and final quarter and semester essays. In order to make it clear to his students that the information conveyed during the workshops was important, workshop attendance was graded. Students had to attend a certain percentage of them to receive an "A", "B", etc. for that subcategory of the individual work grade. In addition, the quality and timeliness of the journal entries associated with the workshop work was graded. Since Gary wanted students to reflect upon what they had learned through the workshops and project work to come to conclusions about the influence of science and technology on human lives, he asked them to write an integrative essay at the end of each academic quarter. The grade for this essay was another subcategory of the individual work grade. Finally, there were a few short, individual projects, such as the Home Inventory project done in the first quarter. These were graded as part of individual work.

Last, but certainly not least, the personal outcomes category was used to grade the extent to which students exhibited "self-directed learning", as well as how well they worked with others and managed time and organization of materials. The grades in this category were similar in some ways to both Roger's work grade and Carol's personal grade. Their projects binders and journal work served as part of the evidence reviewed for this category. In addition, Gary and Karen kept observation notes for each student. However, these notes were fairly limited in scope. For many students they had taken none at all. Thus, the personal outcomes grade was largely determined by their memories of classroom behavior as well as the body of work each student generated.

Gary's course and assessment infrastructure designs were meant to provide students with both freedom and control over their learning and concrete consequences for the choices they made in a non-coercive fashion. The holistic grading scheme and associated grading conversations were meant to turn assessment into a dialog between Gary and his students. While the final grading process might be somewhat "mushy", the rest of his assessment infrastructure was not. The requirements for binders, final products, and individual work were quite specific and the evaluation criteria for assessing them matched those specifications. Moreover, Gary was very proactive in his patterns of conferring with students about their work—both the concrete aspects of it, such as the milestones and final products, as well as their working habits and behaviors. Since teaching his students *how* to learn in an open environment like that of his class was a key goal of his, the students work habits were just as important as the completed work itself and received significant attention in his grading scheme.

Summary

A review of all the assessment infrastructures crafted in all the courses shows that despite key differences in the way each teacher conceived of and designed them, there were several important commonalties. First, tests, homework assignments (e.g. reaction papers and answers to questions on readings), lab write-ups and reflective essays were all used to assess students' knowledge and understanding of core concepts learned by the entire student body as a whole, largely through traditional means (e.g. lectures, labs, discussions, assigned readings). These means were *not* used to assess the students' project work or what they had learned from their projects. Second, the kinds of artifacts and behaviors which were evaluated in order to assess and support students' project work were fairly consistent across the three classrooms. All three required the students to produce a final product (e.g., papers, presentations, videos, etc.) which demonstrated what was learned in the course of the project. All three required that some form of intermediate products (milestones, drafts, ancillary reports) be submitted by the students during the course of the project. All three kept records and graded classroom behaviors related to projects, including use of class time, ability to collaborate with others, and/or ability to work with outside adult experts. Finally, all three found that face-to-face conferences with students were critical to their project success and so established routines for interacting with the project teams. In the next section, the way in which each of these assessment components supports the particular challenges of project science is explored in depth.

MEETING THE CHALLENGES OF PROJECT-BASED SCIENCE PEDAGOGY

While technology was a strong component in the instructional designs of all three teachers, the nature of project-based pedagogy was clearly the more dominant concern for them as they structured classroom assessment infrastructures to support their new visions of teaching and learning. There are four aspects to the nature of project activity which differ significantly from traditional science courses taught primarily through lectures, textbooks, and labs. Student work on science projects tends to be (1) long-term and open-ended, (2) resource intensive (3) cooperative and/or collaborative and (4) iterative. Each of these aspects of the nature of project work resulted in a cluster of support and evaluation challenges which the teachers studied here attempted to meet in part through their assessment infrastructure designs. Below, each aspect of the nature of project science, its challenges, and the assessment infrastructure designs the teachers used to meet them, are discussed.

The long-term, open-ended nature of project work

In these classrooms, project work was considered long-term and open-ended when students had room to shape and/or choose their own learning goals, as well as use class time in their own way over periods of weeks or months. Giving students such control over the focus of their work and management of their time created several challenges for the teachers involved in this study. They discovered that students new to the complex practice of science often did not know how to craft a question or issue to investigate that was both feasible and interesting given their time and resource constraints. Moreover, the teachers reported that many students had difficulty determining the logical subtasks required to accomplish their projects, estimating how long each would take and managing their time so as to finish by the deadline. In general, the more open-ended the project cycles were, the more difficult it was for the teachers to provide a general set of expectations for the work or guidelines that students could use to direct their efforts. The teachers managed these challenges by providing students with structural supports for their projects—either by creating less open-ended (i.e. more teacher-directed) project cycles or through the design of assessment infrastructures. Key assessment infrastructure components used to mediate the challenges associated with the long-term, open-ended nature of project science were intermediate products (including milestones, drafts, and ancillary reports), face-to-face conferences with students, and records of students’ use of class time.

Providing structure

One way to guide students through long-term, open-ended project work is to provide more structure to their activities. The teachers in this study created structure by either designing more of the project activity themselves (e.g. specifying guidelines or restrictions for the topic or issue to be studied, the procedures used to explore it, the format of the final product, and/or the kind of data collection and analysis done), or by requiring and assessing intermediate products from the students as they worked. The first strategy enabled the teacher to provide more specific directions for doing the work and more specific guidelines for the final product. While such specifications are not part of the assessment infrastructure per se, they do constitute a publicly understood set of expectations which are linked to the assessment infrastructure. For example, the specific guidelines given for the final product often became the dimensions along which that final product was evaluated and graded. Guidelines for how to do the work sometimes became reified as intermediate products. Carol and Gary both made substantive use of this strategy and had several project cycles within their courses that were fairly structured ones. For these projects, their directions for work and guidelines for final products were often quite specific.

In the second strategy, teachers provided structure by assigning intermediate products and associated deadlines. Intermediate products in the form of project milestones and drafts of final products helped students break down the work of accomplishing their projects into logical subtasks. Moreover, by providing intermediate deadlines, they helped students pace their work. Finally, as discussed further in the next section, they provided an occasion for students to confer with their teachers about their project progress, either in preparation for meeting the deadline or in response to the teachers comments on their intermediate work.

Roger and Gary both made significant use of this technique. Each required sets of milestones for the projects in their courses each of which were either pieces of the final product (e.g. Roger’s milestone for a written summary of the background information on their topic became the introduction to the team’s paper) or were the natural outcome of work necessary to complete the project (e.g. Gary’s milestone of research notes from 10 sources by a particular deadline). However, with the exception of paper drafts, Carol’s intermediate products usually took the form of ancillary reports (e.g. progress reports).

Interviews with students showed that in general they approved of and appreciated the milestone and draft deadlines. Since they were directly relevant to accomplishing their goals, the students saw them as work worth doing. As one of Roger’s students explained:

> He gives you a project, or a guideline. Let’s say, in one week you have to hand in your partners. And then the week after that you have to know your main subject. You have to get some background information on it. So each time there’s a deadline. You hand in your partners name, your topic a week later, and two weeks later you hand in your background information. Each one of those is graded—five [or] ten points. And so that’s basically like the system. You just follow it, you turn stuff in when its needed to be turned in and you’ll get the points. And after the background comes like, your data and your data analysis and your conclusion. And, you do this one step at a time and you hand all the things in, get graded and then in the end it comes together as a big project. (James, interview, February 7, 1996).

As will be discussed in the next section, the students did not see the ancillary reports as equally valuable.

Providing individualized support

The structure provided through either less open-ended projects or the assignment of intermediate products gave students general guidelines for doing and pacing their work. However, such structural supports were necessarily somewhat generic in nature. During even the most structured projects seen in these three classrooms, each project team had support needs *specific* to their project. For example, the assessment of risk due to food additives explored by students during Carol's Food Additives project, was based on each individual's diet. Differences in their diets resulted in the need to explore risks associated with different food additives, and information on each was not necessarily found through the same resources. Moreover, some students studied particular kinds of risks, such as foods and food additives linked to migraine headaches. Thus, while the structured nature of this project cycle provided general guidelines for their work, the students still needed individualized assistance with their specific problems.

In more open-ended projects, where students had significant control over the topic explored and the design of their study (i.e., the projects in Roger's class, or those towards the end of the year in Carol's and Gary's), the need for assistance to be tailored to the specifics of each team's project was even greater. Students in these cases not only needed assistance to solve problems they encountered along the way, they also needed help choosing a feasible and interesting question to explore and designing and effective study for answering that question. All three teachers provided individualized support to the project teams through the assignment and evaluation of intermediate products as well as face-to-face conferences with the students.

In addition to milestones and drafts, ancillary reports from students were used to evaluate how each project team was proceeding and provide students with feedback on their work in Carol and Gary's classes. Usually the teachers made significant written comments on the milestone and draft submissions. As mentioned earlier, the deadlines associated with milestones, as well as comments made on the completed versions, often provided an impetus for students and teachers to confer with one another. Ancillary reports, on the other hand, were often only discussed verbally. The progress reports Carol's students submitted to her were tied to her "touch-base" days and the progress reports and task logs Gary's students kept in their binders were tied to their binder checks.

Student interviews indicated that the ancillary reports were a less reliable reflection of the project team's work than the milestones and drafts and therefore a less useful focal point for conversations between teachers and students on how to proceed. Essentially, the students saw the progress reports Gary and Carol required as ancillary to their work. They did not have a direct connection to the students' final goals and so they spent little time completing them, sometimes lying about, or at least "spinning" the information they provided in them. One of Carol's students explained that he was able postpone much of the work for his project, "By saying that I had done stuff that I hadn't. So that was, so she does check. But it's not like turning something in or actually showing her a list of twenty additives [for the Food Additives project] and the research you've done, you know, stuff like that," (Jason, interview, February 6, 1996). As a result, the purpose of the "touch base days"—to provide feedback, course corrections, and advice specific to each student's or team's project—could easily be circumvented by students. Only later might they realize the importance of such advice, as Jason did, "I definitely put the project off which kind of-- I think I could have done a lot better if I didn't because she didn't get a chance to look over it and while I'm doing it at the last second, I was kind of unclear. I had a few questions as far as what she was looking for in the paper," (Jason, interview, February 6, 1996).

While the intermediate products served as an impetus for tailored feedback through face-to-face conversations, they were not the only source. During the days devoted to project work, students in all three classes could and did approach their teachers for assistance and advice. Roger's style of interacting with students was largely reactive. Polman (1997) found that the majority of the conferences Roger had with students were initiated by the students. In contrast, Gary was very proactive, circulating constantly among the students and asking them about their progress. He often spoke briefly with members of nearly every team in the class at least once during their double period session and spent significant time with between one and three groups. Carol's conferring patterns lay somewhere between these two. She worried about "butting into" students work and so was somewhat, though not completely, reactive in her interactions. One student described her routines as, "in the middle between like up in your face, like, you know, and totally, just, like, you always have to go to her. She's more toward the you have to go to her. But she's in the middle, which is, like, perfect, I think," (Jason, interview, April 18, 1996). Her "touch-base" days countered any reluctance to invade students territory by creating a formal event that ensured she met with every group and provided them feedback.

Students in all three classes mentioned in interviews how critical these face-to-face conversations were to completing a successful project. For example, one or Roger's students, Papi, pointed out that part of the reason they were so important was that talking to Roger was the only way to get *specific* advice about your project:

> When he knows what's going on, he can give you advice you can use, and not just advice that's so general that you can apply it to anything you do, like any kind of experiment. If you keep him informed of what's going on, he can help you pinpoint your problem and help you find, if not a solution, a way of coming to a conclusion for it. I mean it seems he knows what he's doing, (Papi, interview, February 8, 1996).

However, getting advice from Roger was not always easy. His attention was often split between various students who needed his help as well as administrative parts of his job that commanded his attention. In addition, his personal style and brand of humor was intimidating to some students. Tweed, who believed that "there's a chemistry involved in getting an 'A'. [...] you have to have some sort of personal one-on-one relationship with the teacher to be able to get an 'A'," felt comfortable interacting with Roger, "Like I'm very cocky in that class and I definitely allow, like there's teasing. There's like, 'Tweed, you're wrong.' 'Well, Roger, you're wrong.' There's definite, like we're comfortable in doing that and that allows, like, honesty," (Tweed, interview, February 16, 1997). At the same time, she worried that other students might not be as comfortable interacting with him as she was and that they might not be getting the help they needed to do their work well or develop the rapport the necessary for doing "A" work:

> I think he's cocky. I really think he's cocky when he's teaching. When someone doesn't understand something and he goes, "What don't you understand?" Like what the hell, you know, they don't understand. Meaning you need to slow down and re-say what you just said or somehow rewrite it so they'll understand, not act like they're idiots. You know? These things don't come easily to people. This is not reading and writing and arithmetic. This is different thinking than most people do on a daily basis (Tweed, interview, February 16, 1997).

The interviews with students in the other two classes exhibited a similar overall pattern of concern—conferring with the teacher(s) is important to project success, but a given students' ability to have a productive conference session with the teacher(s) affected by the interaction between the teacher's conferring style and personality as well as the student's personality and willingness to heed the teacher's advice.

In sum, conversations were critical to providing students the assistance they needed to design feasible projects and solve problems along the way. The more open-ended the project, the more critical such conversations became and the more likely it was that personality conflicts between teachers and students could lead to project failure through lack of sufficiently specific feedback and guidance. Intermediate products such as milestones and drafts were successful both as indicators of student progress and as an impetus for these face-to-face conversations. Ancillary reports were less useful for these purposes as the students typically did not represent their activities well within them.

Encouraging good work habits

Teachers used their conversations with students not only to provide feedback and advice to students but also to encourage them to use their time well. In order for students to successfully complete their projects, they had to make effective use of class time devoted to project work, which in these three classrooms was a significant amount of time.[6] Since each project team usually had different goals and therefore a different set of tasks to complete, the class time teachers provided for project work was typically very open and unstructured. Each team directed their own activities, while the teacher served as a consultant and supervisor. Since the projects were typically long term ones and the amount of available class time to work on them was large, students interviewed in all three classes noted that it was easy to procrastinate. As one of Roger's students said:

> The trouble that many people had, it's obvious, is that they felt at first, being the first project and they didn't have a teacher on their back stressing them to do their assignments, that they could blow off the class. They realized a couple nights before their project that, you know, "I got a lot of work to do!" and they crammed it (Papi, interview, February 8, 1996).

The three teachers used a combination of intermediate deadlines, face-to-face conversations with students and records of students' use of class time in order to encourage and reward strong work habits.

Because Roger was generally reactive in his interaction style with students, he neither prodded nor cajoled them to work. Instead, he made the consequences for wasting time clear—students would loose points on their daily work grade for inadequate use of class time, loose points for milestones they turned in late and lose points on their final paper if they failed to put in the effort needed to successfully and convincingly complete their research. As one of his students, Tweed put it, Roger's class was "an easy D". Roger would allow students to do nothing if they chose, but they paid a price for that choice.

Carol and Gary were less laissez faire. Carol mentioned in one interview, that even with all the project structure she provided her students, they still had difficulties managing their time, "The

[6]Daily logs kept by each teacher in the 1994-95 school year indicated that 45% of the class time in Carol's class, 48% of the class time in Gary's class and 82% of the class time in Roger's class was devoted to project work.

ones that failed there were people who couldn't manage their time in class. [...] I set the structure, and they still couldn't handle it," (C. Patterson, interview, March 10, 1994). As mentioned earlier, both she and Gary were more proactive in their interactions with students. In addition to their regularly scheduled meetings with students, both circulated regularly asking students about their progress and in Gary's case, explicitly commenting on how they were choosing to use their time. They both made it clear through their presence and oversight that class time was for working.

Like Roger, they both felt the need to communicate the importance of using class time well by basing a portion of each student's grade on it. However, in Carol's case this grade made up only a very small portion of the students' grades (less than 5% each semester) and in both cases was based on a less systematic appraisal of student work habits than Roger's. Both assessed students' use of class time in combination with several other classroom behaviors, such as perseverance in the face of project set backs and ability to work with others, and were based on each teacher's memory of class room events supported by some opportunistically taken notes.

While all three teachers felt the need to grade students on their use of class time, it was the portion of their assessment infrastructures with which both they and their students were the least satisfied. In fact, when asked what they would change about the grading in the course, the component most frequently mentioned by Roger's students (n=48) was his work grade (17%). Students complained to Roger that he couldn't possible keep tabs on 20-30 students in each class (some of whom on any given day were in the library) and say for certain who was on or not on task. He agreed with them adding, "When they're on the computer and they're doing email, is it project related email or are they writing to their friends? If they're on Netscape, are they looking for, you know, music and lyrics or are they actually looking for volcanoes and earthquakes?" (R. Wolfe, interview, March 7, 1996). However, although he tried using student records of class time use, he was skeptical about their reliability, "I have very little faith in student self evaluations. [..] Because I don't think they're honest. And it may not, and it may not be willful deceit. They just may not have a clue as to what they're actually doing," (R. Wolfe, interview, March 7, 1996). Carol had similar concerns. She described one project cycle in which she asked students to keep logs about their project work, "I had them use a log. I know that a lot of what I saw was bogus on the logs because I know, and I was there every day and I saw that most of them weren't doing anything," (C. Patterson, interview, March 10, 1994).

Gary, on the other hand, who asked students to report how hard they worked and the number of hours they dedicated to their project work in their weekly progress reports, did not seem particularly concerned about the veracity of students' self-reports. He may have been less worried about them because they were only on of several indicators he considered when holistically assigning students personal outcomes grade and therefore did not *have* to credit them in any particular way. However, he did receive complaints from students who claimed they did most of their work at home and therefore felt Gary could not *see* their use of time or other behaviors that he considered as part of his Personal Outcomes grade.

Of the three teachers, Carol received the fewest complains about her grades on students work habits and other classroom behaviors. The complaints in her class were probably small both because the these assessments constituted a very small portion of the students' final quarter and semester grades and because students (rightly) perceived that it was largely used to their benefit. In other words, Carol did not so much penalize students for poor work habits as reward them for good ones.

Interviews with students in all three classrooms indicated that in general they believed effort and dedication put into a project was inconsequential if the end result was a solid final product. However, if the final product were a poor one, *despite* diligent effort on the part of the project team to overcome any problems and extenuating circumstances they might encounter, *then* students felt their perseverance and use of class time should be considered in the final grade.

The resource intensive nature of project work

Students often had difficulties obtaining and using the resources needed to complete the projects they proposed. Projects could and did fail because students were unable to fulfill their resource needs. None of the teachers wanted to fail students because of unforseeable resource roadblocks, particularly when those students made genuine efforts to acquire what they needed. To avoid this situation, the teachers needed to make certain their project cycle and assessment infrastructure designs helped them manage two major challenges associated with the resource intensive nature of project work (1) student mobility and (2) resource conflicts and vacuums. Components of the assessment infrastructure used to handle resource problems included: records of students use of class time and movement; intermediate products; and face-to-face conferences.

Providing students with mobility

The first of these challenges is students' need to be mobile in order to access resources. All three teachers in this study wanted to allow students the freedom to spend class time finding the resources they needed. However, when those needs necessitated access to sources not available within the classroom or school walls, the teachers had to make decisions about how much mobility they were going to allow their students, under what conditions they would allow that mobility, and with what consequences if the freedom was abused. All three allowed their students to leave the classroom (and in Gary's case, the school building) to get the resources they needed. While giving the students freedom to leave the classroom made it less likely that their projects would fail because they did not have adequate access to resources, it created a “student supervision” problem for the teachers that each handled differently depending on the school rules for student mobility and their own concerns about student supervision.

Both Gary and Roger set up consequences for students who abused their freedom of movement. Except in extreme cases, these consequences were realized through the students’ grades rather than other disciplinary action. Lakeside had a pass system that allowed students to leave their assigned classroom and travel to other parts of the school. Gary worked with the administration to create a special research pass for his students that extended this system to include destinations outside the school building. These passes needed to be signed by an adult at their destination point and returned to Gary or Karen. A missing research pass was equivalent to an unexcused absence. In Gary’s course, each unexcused absence resulted in a half grade deduction (e.g. from a B to a B-) from the students’ quarter grade. There was no pass system at all in Edgewater and students were allowed hallway passage at any time. Roger, worried that not all his students would go where they said they would go, performed periodic “library checks” during his class. Students who had been released to the library but could not be found there when he checked received a zero for that day’s work grade. Students in both classes who seriously abused their mobility privileges had them revoked. Carol seemed less concerned that her students would abuse the privilege and chose to simply trust them.

Avoiding failure due to resource conflicts or vacuums

Determining the resources they needed to complete their projects, as well as where to find them, was often as difficult a challenge for the students as formulating a researchable question. As Roger once said, a large part of “helping kids do projects is just knowing where the datasets are that can be used,” (R. Wolfe, interview, March 10, 1994). Moreover, data and information which both teachers and students might have presumed likely to exist (e.g., records of geyser eruptions at Yellowstone National Park), might be difficult to obtain, or if found in formats that are difficult to use and interpret (e.g. records of astronomical data retrieved from NASA). Finally it was not unusual, even in completely open-ended project cycles, for students to be chasing, and in some cases hoarding, the same few key resources. For example, at Edgewater a book entitled *Volcanoes of the World*, which contained a significant amount of data on volcanic activity around the world over a large span of time, was a key resource for any team doing a project related to volcanoes. When it was discovered missing from the school library, there was a mild crisis among Roger’s students as those teams studying the very popular topic of volcanoes scrambled to find alternative sources of data. At the very least, access to the classroom computers or other technologies to support their work could be a source of conflict. As a result of these potential difficulties, all three teachers had to mediate situations in which the students were simply unable to acquire the resources needed to complete their projects.

The teachers tried to minimize the likelihood of complete project failure due to resource difficulties in several ways. First, as already discussed, they allowed students the freedom to leave the class room in order to obtain what they needed. Second, they encouraged students to use the internet as a source of primary data they might be unable to collect themselves (e.g. records of earthquake activity) and information not readily available locally (e.g. legislation on the environment currently under congressional review). Third, they sometimes used scheduling to ensure equal access to the computers and technology in the classroom or elsewhere in the school. Finally, they used a combination of intermediate products and face-to-face conferences to determine which students were having resource difficulties early enough in the project cycle that a course correction could be planned before the project met with disaster.

When these strategies were insufficient, as they sometimes were, the teachers relied on their grades of students work habits, and some creative grading on their final products, to compensate. Gary’s personal outcomes and binder grades and Roger’s work grade were both a significant proportion of each student’s grade and could thus temper some resource problems. Similarly, Carol hoped that their opportunities to show what they had learned in the course through other means (e.g. tests, homework, and labs) could balanced out a poor project. In addition, all three defined subsections of the final product and depending on the extent of the resource difficulty, the student might still do well on some of them. Carol and Roger who assigned specific point values on each of those sections, did not have firm heuristics for how to determine what kinds of problems led to what kinds of reduction in points. As Carol said, “I'll say this many points for this particular section of the paper. What does that mean? I'll take off three points. Well what does that mean? It doesn't--It means nothing,” (C. Patterson, interview, March 22, 1996). As a result, they could compensate for difficulties due to resource failures outside the students’ control through the way they assigned grades to each portion of the final product:

> I got a feeling for just the overall quality of the project and I thought, “Well, [...] it feels more to me like a B than a C, because of all the other things that he did which were not required but he's trying [...] Okay.” Maybe if he had not done anything

> extra special, then for what he did he would have gotten a C. [...] And the points, I can't tell you exactly about the points, but I'll, I'll work it out some way so that he ends up getting a B (C. Patterson, interview, March 22, 1996).

The cooperative or collaborative nature of project work

Often the work in these three classrooms was done collaboratively or cooperatively. Students worked on project teams and often had to interact with adults outside the classroom who were either advisors (telementors), sources of data and information, or objects of study. Moreover, as part of their attempts to establish a common set of shared expectations for good work, they each tried to engage the students in discussing and evaluating one another's work. This kind of dependence others for the success of their work was generally uncomfortable for at least a portion of the students in each class and sometimes created difficulties similar to those created by the resource intensive nature of project work. In particular, they worried that their project team might successfully complete their project, but that the labor for doing so might be unfairly distributed among the team members. The "free riders" would then receive the same grade as those students who had contributed significant effort to the project, a situation deemed unfair to the teachers as well as the students. They also worried that interpersonal problems within the group would cause such dysfunction that the final product would be either incomplete or inadequate.

The teachers used their assessment infrastructure to compensate for these difficulties using strategies similar to those they used for handling resource complications. They relied on the use of face-to-face conferences both to reveal when the students were having interpersonal difficulties and to help them resolve those difficulties before they became fatal to the project's success. Second they relied on their grades of students work habits to reward students who did more than their fair share of the project work. In order to convey to students that mature and productive interpersonal relations were a goal of the course, Carol included a rating on "group cohesiveness" as part of the grade she gave on the final products of the group projects. Likewise, Gary graded students ability to work with or lead team members and Roger asked the students to rate one another's contributions to the project work. In order to encourage mature relations with outside adults, both Carol and Roger experimented with, though never really enforced, grades on students interactions with their mentors. Finally, all three teachers had the students evaluate the work of the other project teams in the class, usually based on their final presentations, in order to establish a common culture of expectations and values around the work and build students capacities for looking at the work critically. However, they only sometimes incorporated these evaluations into their final grades and when they did so, it was a minor component.

The iterative nature of project work

The work the teachers were asking the students to accomplish in each of these classes was often very different from the work they had been asked to do in other courses. When asked on surveys to rate on a scale of zero to ten how different this class was from other science classes, the students in all three responded that it was very different. (See table 3 below.) Moreover, they found the projects were different from those they had done in other science classes as well. (See table 4 below.) Thus, the students could be expected to have difficulties accomplishing the projects at first. They needed practice. All three teachers expected and hoped that the students would get better at doing them over the course of the year. However, they never directly assessed students' improvement. They did not have an explicit improvement component to their grade (though it may have been captured in Gary's personal outcomes grade or Carol's participation grade), nor did they explicitly change the criteria for evaluation over the course of the year so that what constituted good work was different in the spring than in the fall. Carol and Gary did reduce the amount scaffolding they provided students through the course of the year, by giving them more control over the project design. Therefore, an "A" earned in the spring was for a more complex task than that done in the fall and might in this way demonstrate student's improved capacity for doing projects.

	Respondents	**Mean**	**Median**	**StDev**
Roger's students	51	1.78	1	1.96
Carol's students	14	3.2	3.5	1.67
Gary's students	45	1.89	2	1.71

Table 3: Difference between this science class and other science classes
(0=completely different, 10=very similar)

	Respondents	**Mean**	**Median**	**StDev**
Roger's students	43	2.74	3	2.30
Carol's students	11	3.54	4	2.38
Gary's students	39	2.85	3	2.42

Table 4: Difference between the projects done in this class and in other science classes
(0=completely different, 10=very similar)

They also expected the students to run into research failures that might require them to redo their work. The school marking periods made it difficult to give students more than eight or ten weeks to work on a project and all projects were typically due at the same time regardless of what challenges individual groups may have faced. As a result, while students often had the chance to rewrite their final papers, they rarely had the chance to redo the research. Rewriting the paper would not help the students' grades much if the research they had done was poor. All three teachers tried to avoid the need for students to repeat their research in the same way they tried to avoid project failures due to resource difficulties and teamwork problems. They assigned intermediate products to help them gage which students were having difficulties conducting their research and through their face-to-face conferences attempted to circumvent the need of students to redo work by trouble-shooting difficulties early on.

Reviewing and grading drafts was a somewhat new experience for these three teachers and they each handled it differently. They each needed to develop criteria for evaluating drafts that recognized their potentially incomplete nature, yet did not abandon all standards for quality. Roger's solution was to never tell his students their first attempt at their paper was a draft. "Because if we tell them its just a draft, they don't take it seriously," he explained (R. Wolfe, interview, June 2, 1995). Instead, he had them turn in the final paper and then after he had graded it, gave them the opportunity to rewrite it if they so desired. He then gave them the higher of the two grades. Carol required drafts of all her students and while she provided them with copious notes and feedback, did not grade them. Gary often had his students continuously turn in drafts until the final product was good enough or he and the students gave up, whichever came first. Students in his class were graded not only on their final product, but also on the drafting process as part of their personal outcomes and binder grades.

Summary

As the discussion above illustrates, many of the assessment infrastructure components used by the teachers were intended to serve more than one need. For example, the intermediate products required were often intended to help students pace their work (long-term, open-ended nature of project work), enable the teacher to make certain the resource requirements of their project were reasonable (resource intensive nature of project work), and make it more likely the teacher could steer the students through research difficulties that might require revisions (iterative nature of project work) they did not have time to do. Notes kept by either teachers and/or students about their work activity were used to help teachers guide students in their decision making, track and guide their use of class time (long-term, open-ended nature of project work) and enable teachers to see how labor was being divided among team mates (cooperative/collaborative nature of project work). Participation or "effort" grades allowed teachers to give credit to students who made reasonable plans, yet ran into unforeseeable and sometimes fatal difficulties (resource intensive and iterative nature of project work). They also made it possible to reward students who helped others with their work, or contributed more than their fair share of the labor on their project (cooperative/collaborative nature of project work). The various pieces of the assessment infrastructures were generally designed to support and reflect one another in this fashion.

CONCLUSION

> "There's a say, I can't remember the exact saying, but-- Assess what you value. You know? And for the student, it's "value what you get assessed on". [...] If they don't know what they're going to be assessed on, they don't know what to value. [...] They need to know what's valuable. [...] Especially because we're aiming for different things in here than in most classes, you know. It's an alien language to begin with," (G. Magi, interview, June 7, 1996).

Gary, described by one of his students as a, "little hippie guy who's like peace and love and harmony," (Guido, interview, May 28, 1996), wanted more than anything for his students to be motivated to learn by their own desires and not by his command. Yet even he realized that students, like most of us, rarely make choices based on simple desire. Like most of us, they have obligations, commitments and enough work to do that it would be difficult to do it all well. So, they make choices. Those choices are based partly on desire, but also partly on expediency. High school teachers, including Gary, know this to be true, so they place value on what matters to them through the currency system they have at their disposal—grades.

Cohen (1988) notes that the success of teachers, like other practitioners who work in human services, is dependent upon the success of their clients, in this case, the students. Their very dependence makes it difficult for them to choose forms of teaching and learning which are complex and highly ambiguous. "When practitioners weigh choices between more and less uncertain objectives, they also weigh how much they are willing to depend on their client's will and skill" (pp. 40–41 of a draft version). Looked at in light of Cohen's statement, the assessment infrastructure design choices of these three teachers can be seen to a certain extent as part of a balancing act intended to reduce the ambiguity inherent in project activity. Thus, the assessment infrastructures created by these teachers were designed not only to evaluate kinds of learning that could not be assessed through other means, but also to shape student activity by communicating what was of value in this new classroom culture, using a language with which the students were already familiar.

The findings outlined in this paper focused largely on the *structural* aspects of classroom assessment infrastructures—what teachers assessed, how they combined those assessments into forming a grade, and what the customs and routines for exchanging assessment-related information and artifacts were—and what effect they had on students' perceptions of those goals. This focus led to three key limitations of the work, all of which concern the *content* that flowed through these structures. First, there was no systematic or deep study of the specific criteria each teacher used when evaluating student work and performance. While general or theoretical heuristics were discussed in interviews, the *actual* reasons behind the teachers' grading decisions on particular student artifacts or activities were not reviewed. Messages about what constituted quality work and performance in each course were conveyed as much by what products and activities received an A as by what products and activities were required. Insights into the messages conveyed to students through such grading decisions are missing from the analysis. Second, there was no systematic analysis of the content of the feedback and direction teachers gave students, neither in the form of written handouts, in the written comments on their work, nor in the course of conversations and meetings with students. Such feedback may be the heart of where teaching and learning actually occurs in these classrooms. While the assignment of intermediate products and the development of routines for meeting with students might provide the opportunity for such teaching and learning to occur, they do not necessarily a guarantee that it *will* occur, nor that it will occur in an substantive fashion. This research is unable to answer questions about which kinds of feedback are most productive for improving student learning and performance. Finally, there was no evaluation of the quality of student work in each of the classes or what they were learning through their project activity. Thus, evaluation of the "effectiveness" of the classroom assessment infrastructures is confined to the perceptions that teachers and students have about how they guided students' priorities, activities, and understanding of teacher expectations, and *not* on their ability to foster quality work or deep understanding.

Despite these limitations, the study suggests several cautions and recommendations for designing assessment infrastructures for project-based science classrooms. First, intermediate products in the form of milestones and drafts were seen as a very useful tool by both students and teachers for guiding students work. They helped students divide the work into its relevant subtasks and pace the accomplishment of those subtasks, particularly with longer and more open-ended project cycles. They provided motivation and time management supports for many students, some of whom were neither highly dedicated to do the work, nor had particularly sharp time management skills. Moreover, the deadlines associated with them created pressure points that encouraged students and teachers to talk about each project team's progress. In contrast, intermediate products in the form of ancillary reports neither helped the students to pace their work, nor provided adequate information to teachers about student progress for the purposes of providing advice and feedback. These findings suggest that teachers implementing projects, especially long-term, open-ended ones, should incorporate intermediate products in the form of milestones and drafts, rather than ancillary reports, into their assessment infrastructures.

Second, conferring with students during the course of their project cycles was crucial—particularly in open-ended cycles—as it was the only way for students to get *specific* assistance for *their* projects. The more open-ended the project, the less likely that teachers could a priori provide guidelines that would be sufficient to the student's task of individual project teams. The need for conferring made interpersonal issues between teachers and students a key consideration. Students who needed help and did not get it, either because they did not ask or because the teacher did not know how to communicate with them, were at a higher risk for project failure. These findings suggest that teachers implementing projects, particularly open-ended ones, need to establish a system for regular meetings with student teams and give serious thought to how they will handle the needs of shy or interpersonally difficult students.

Third, while students usually had the opportunity to rewrite their final reports, the periodicity of the grade reporting system in each school did not usually allow for sufficient time to redo the research which led to their final reports. Typically, the time frame for the project cycles did not allow for a failed project to be redone. Therefore, teachers had to build in assessment mechanisms that would allow them to catch potential project failures before they happened (e.g., progress reports, conferences, milestones). They also felt the need to include compensatory structures in their assessment infrastructure that rewarded students for honest scientific endeavors that nontheless did not result in solid project findings. Teachers implementing project-based science in schools with similar grading and marking routines would need to make similar provisions.

Finally, all three teachers incorporated the evaluation of students' work habits and/or attitudes into their assessment infrastructures to encourage appropriate classroom behavior, compensate students who made reasonable plans and yet ran into unforeseen difficulties they were not able to overcome, and reward students who contributed more than their fair share of the labor on the project. Despite the fact that both teachers and students felt the need for the assessment infrastructure to take classroom behavior into account, it was one of the most frequently mentioned areas of dissatisfaction among teachers and students. Carol Patterson's students seemed to approve of her technique of minimizing the role of such assessments and using them largely for compensatory purposes. Gary's and Roger's techniques, on the other hand, which generated a lot of discussion about appropriate behavior, were uncomfortable and sometimes frustrating for at least some of the students. It is not clear whether any of the techniques used by the three teachers actually helped students behave more appropriately in class. The findings from this study do not suggest a clear solution to this tension. However, they do indicate that the assessment issues underlying it are important ones for project-based science teachers to consider.

In sum, changing the practices of teaching and learning, as well as the values which define classroom cultures to support inquiry, is not an easy task. These teachers sometimes found it to be a formidable one. Often things did not go well, especially in their early attempts. Each teacher experienced one to two years of radical reconstruction before settling down to small scale tinkering. In that time, students whined a lot, often about their grades. Some even dropped the course. Project cycles fizzled, and particular project teams failed outright. Sometimes, the teachers would look at the work their students had produced after much time and effort and think to themselves, as Carol did, "I just feel sick because this is what happened in four weeks time, sometimes five weeks time [...] and I have nothing tangible to look at and say, 'Boy, this has really be time well spent,'" (C. Patterson, interview, July 25, 1996).

Given all the challenges associated with establishing a project-based science classroom and its supporting assessment infrastructure, one might reasonably ask, "Why pursue this challenge at all?" As mentioned in the introduction, the response of educational reformers is that traditional forms of science education do not develop adequate forms of scientific understanding or practice. Concerns about this inadequacy have led them to call for learning activities which are inquiry-based and engage students with the complexities of scientific knowledge and practice. Project-based science fits this profile. But, perhaps the words of one of Roger's students will be more persuasive:

> Because you're going to do it for the rest of your life! I mean all this other shit that you're doing in school you're never going to use again. You're never going to have to write a five paragraph essay and have someone critique it. You're going to have to have to ask yourself questions and find out the information. It's very realistic. It's just a pain in the butt after three years of specific rules and regulations to be completely broken of that. (Tweed, interview, February 16, 1996).

For those who wish to make the attempt, this report has described the assessment infrastructures designed by three mature project based science teachers. Each had very different instructional goals, yet each created infrastructures based on the same basic set of assessment components. Each made design involved trade-offs, and each was imperfect. Nonetheless all three made very real progress toward breaking students of the "specific rules and regulations" habit Tweed describes. This is an essential first step toward helping students learn how to engage in the practices of complex domains such as science.

References

American Association for the Advancement of Science (1990). *Science for all American's.* New York: Oxford University Press.

American Association for the Advancement of Science (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

Airasian, P. W., & Jones, A. M. (1993). The teacher as applied measurer: Realities of classroom measurement and assessment. *Applied Measurement in Education,* 6(3), 241-254.

Austin, S. & McCann, R. (1992, April). "Here's another arbitrary grade for your collection": A statewide study of grading policies. Paper presented at the Annual Meeting of the American Educational Research Association, April 23, 1992, San Francisco, CA.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.

Borko, H. & Elliott, R. (1998, November). *Tensions between competing pedagogical and accountability commitments for exemplary teachers of mathematics in Kentucky* (CSE Technical Report #495). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California-Los Angeles.

Borko, H., Flory, M. & Cumbo, K. (1993). Teachers' ideas and practices about assessment and instruction. Paper presented at the Annual Meeting of the American Educational Research Association, April 1993, Atlanta, GA.

Briscoe, C. (1994). Making the grade: Perspectives on a teacher's assessment practices. *Mid-Western Educational Researcher, 7*(4), 14-16, 21-25.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Carey, S., Evans, R., Honda, M., Jay, E., and Unger, C. (1989). "An experiment is when you try it and see if it works": a study of grade 77 students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11*(special issue), 514-529.

Cizek, G. J., Rachor, R. E. & Fitzgerald, S. (1995, April). Further investigation of teachers' assessment practices. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

D'Amico, L. (in progress). *The role of classroom assessment infrastructures in crafting project based science classrooms.* Unpublished doctoral dissertation. Northwestern University.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York, NY: Teachers College Press.

Edelson, D. C., Gordin, D. N. & Pea, R. D. (in press). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences*.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching*, (Third ed., pp. 119-161). New York: MacMillan Publishing Company.

Fishman, B. J., & D'Amico, L. M. (1994, June 25-30, 1994). *Which way will the wind blow? Networked computer tools for studying the weather.* Paper presented at the World Conference on Educational Multimedia and Hypermedia (ED-MEDIA 94), Vancouver, BC, Canada.

Freedman, S. W. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment*, 1(1), 27-52.

Gordin, D., Polman, J. L., & Pea, R. D. (1994). The Climate Visualizer: Sense-making through scientific visualization. *Journal of Science Education and Technology, 3*(4), 203-226.

Guskey, T. R. (Ed.). (1994). *High Stakes Performance Assessment: Perspectives on Kentucky's Educational Reform*. Thousand Oaks, CA: Corwin Press, Inc.

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal of Research in Science Teaching, 24*(4), 291-307.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). A practical guide to alternative assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Kulm, G., & Malcom, S. M. (Eds.). (1991). *Science assessment in the service of reform*. Washington, D.C.: American Association for the Advancement of Science.

Lane, S., Parke, C. & Moskal, B. (1992,). *Principles for developing performance assessments*. Paper presented at the 1992 Annual Meeting of the American Educational Research Association.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.

Lave, J. (1990). The culture of acquisition and the practice of understanding. In R. S. J. Stigler, & G. Herdt (Ed.), *Cultural psychology: Essays on comparative human development*, (pp. 309-328). New York: Cambridge University Press.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge: Cambridge University Press.

Lederman, N. G. (1992). Students' and teachers' conceptions of science: A review of the research. *Journal of research in science teaching, 28*(4), 331-359.

Linn, M. C., Songer, N. B., & Lewis, E. L. (1991). Overview: Students' models and epistemologies of science. *Journal of Research in Science Teaching, 28*(9), 729-732.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model.* Alexandria, VA: Association for Supervision and Curriculum Development.

Messick, S. (1994). *Alternative modes of assessment, uniform standards of validity* (Research Report): Educational Testing Service, Princeton, NJ.

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge*. Madison, WI: University of Wisconsin Press.

New Standards Project. (1997). New Standards Reference Exams in English Language Arts and Mathematics. Harcourt Brace.

O'Neil, J. (1993). On the New Standards Project: A conversation with Lauren Resnick and Warren Simmons. *Educational Leadership*, 50(5), 17-21.

O'Neill, D. K., & Gomez, L. M. (1994, June 25-30, 1994). *The Collaboratory Notebook: A distributed knowledge-building environment for project learning.* Paper presented at the World Conference on Educational Multimedia and Hypermedia (ED-MEDIA 94), Vancouver, BC, Canada.

O'Neill D. K. (1997, March). Bluffing their way into science: Analyzing students' appropriation of the Research Article genre. Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL, U.S.A.

O'Neill, D. K. (1998). *Engaging science practice through science practitioners: Design experiments in K-12 telementoring*. Unpublished doctoral dissertation. Northwestern University, Evanston, IL.

O'Neill, D. K. and Gomez, L. M. (1998). Sustaining mentoring relationships on-line. In Greenberg, S. and Neuwirth, C. (eds.), *Proceedings of CSCW 98: ACM conference on computer-supported cooperative work*. November 14-18, 1998. New York: Association for Computing Machinery.

Pea, R. D. (1993). The collaborative visualization project. *Communications of the ACM*, *36*(5), 60-63.

Pea, R. D., & Gomez, L. M. (1992). Distributed multimedia learning environments: Why and how? *Interactive Learning Environments, 2*(2), 73-109.

Pea, R. D., Gomez, L. M., & Edelson, D. C. (1995). Science education as a driver of cyberspace technology development. Paper presented at the INET '95.

Polman, J. L. (1997). *Guiding science expeditions: The design of a learning environment for project-based science.* Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage Publications.

QSR. (1997). NUD*IST (Version 4.0). Thousand Oaks, CA: Scolari.

Resnick, L. B., & Resnick, D. P. (1991). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement and instruction . Boston, MA: Kluwer.

Ruopp, R., Gal, S., Drayton, B., & Pfister, M. (Eds.). (1993). *LabNet: Toward a Community of Practice*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Saxe, G. B., Franke, M. L., Gearhart, M. Howard, S. B. & Crockett, M. (1997, December). *Teachers' shifting assessment practices in the context of educational reform in mathematics* (CSE Technical Report #471). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California-Los Angeles.

Shavelson, R. J., Baxter, G. P. & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher*(May), 22-27.

Shavelson, R. J. & Ruis-Primo, M. A. (1998). *On the assessment of science achievement: Conceptual underpinnings for the design of performance assessments—Report of year 2 activities* (CSE Technical Report #491). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California-Los Angeles.

Sheingold, K., Heller, Joan I., & Paulukonis, Susan T. (1995). Actively Seeking Evidence: Teacher Change through Assessment Development (94-04): Center for Performance Assessment, Educational Testing Service.

Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1995). Effects of introducing classroom performance assessments on student learning . Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California-Los Angeles.

Stecher, B. M., Barron, S., Kaganoff, T. & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing* (CSE Technical Report #482). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California-Los Angeles.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 2771-286.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8(*2), 5-14.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. New York: Cambridge University Press.

Vickers, M. (1996). New directions in the assessment of high school achievement: Cross national perspectives. In L. B. Resnick & J. G. Wirt (Eds.), *Linking school and work: Roles for standards and assessment* (pp. 341-380). San Francisco, CA: Jossey-Bass.

Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.

Wiggins, G. (1993). Assessment: Authenticity, context and validity. *Phi Delta Kappan* (November), 200-214.

Wolf, D. P. (1989). Portfolio assessment: Sampling student work. *Educational Leadership*(April), 35-39.

The history of pedagogy as a science is briefly viewed in many educational humanitarian programs. It is possible to familiarize with it independently, having studied thematic editions, textbooks, articles. The topic is quite interesting, because pedagogy was born in ancient times and has been actively improved since then, although the development path is rather uneven. Consider what is at stake and what kind of discipline.Â This ancient Greek figure is honored for the founder of science. The information that he taught dialogue, controversy, set himself the goal to form the ability to think in the audience, obeying logic, has reached our days.Â The approach is based on questions and answers through which learning to think is realized, obeying logic. Antiquity: the development of ideas. Project-based learning (PBL) is a student-centered pedagogy that involves a dynamic classroom approach in which it is believed that students acquire a deeper knowledge through active exploration of real-world challenges and problems. Students learn about a subject by working for an extended period of time to investigate and respond to a complex question, challenge, or problem. It is a style of active learning and inquiry-based learning. PBL contrasts with paper-based, rote memorization, or teacher-led Science Teaching Science Teacher Content Knowledge Pedagogical Content Knowledge Heat Energy.Â American Association for the Advancement of Science. (1989). Science for all Americans: A project 2061 reporton literacygoals in science. mathematics. and technology, Washington, D.C.. AAAS.Google Scholar. Anderson, C. W. & Smith, E. L. (1987).Â Verbal explanations given by science teachers: Their nature and implications, Journal of Research in Science Teaching. 29, 361â€“374.Google Scholar. Dobey, D. C., & Schafer, L. E. (1984).