Science talk as a way to assess student transfer and learning:
Implications for formative assessment.
Jackie Gray, Paul Camp, Jennifer Holbrook and Janet Kolodner

Paper to be presented at the Meetings of the American Educational Research Association,
Seattle, WA. (2001, April) .

Our project, Learning by Design™ (LBD™) attempts, like other cognitive-science-derived
approaches to science education, to provide a learning environment that will foster the
development of science knowledge and "habits of mind" for becoming self-regulated life-long
learners.   Our paper, *Problem-Based Learning Meets Case-Based Reasoning in the Middle-*
School Science Classroom: Putting a Learning-by-Design Curriculum into Practice (Kolodner,
J. L., Crismond, D., Fasse, B., Gray, J., Holbrook, J. & Puntambekar, S.,under review), presents
the theoretical foundations behind LBD™, the practical issues we had to address to make LBD™
efficacious, the LBD™ approach that has arisen from those considerations, an implementation of
a sample unit, the trends we are seeing in our pilot and field tests, the lessons we've learned
about making a learner-centered design-based classroom work, and suggestions for others
involved in designing design-based and project-based curricula.

Through its redundant opportunities for designing, testing, and reflecting on constructed
artifacts that work according to scientific principles, students in an LBD environment learn to
practice scientific inquiry and learn important science concepts within a cultural community of
learners that value such practice and learning.   A major focus of our work in Learning By
Design™ implementations over the last several years has been to identify factors critical to
making inquiry work in real classrooms. We are measuring the effect of our implementations in
a large mixed metropolitan/suburban area by assessing student learning. Assessment criteria and
rubrics for assessment can provide an important component of scaffolding. This not only sets
expectations for the students, but it also allows students to set reasonable short-term aims for
themselves, especially if rubrics for assessment make clear the expected developmental
progressions in acquisition of knowledge and/or complex skills.

In this paper, we will address the challenge posed in our assessment efforts to
document students' abilities to collaborate and use scientific reasoning while working on

a complex problem. We focus on the performance assessments we've adapted and the evidence we are documenting as evidence for transfer.

Our approach to program evaluation (or, to assessment) is distinguished by (i) our focus on assessing and scaffolding science and teamwork practices (others focus on assessing knowledge), (ii) our development of resources that scaffold students' performance and learning in a project-based classroom and at the same time scaffold teachers as they develop facilitation skills, and (iii) our design of assessment and scaffolding tools in collaboration with teachers and students so that we know we are addressing real issues in actual classrooms.   Our aim in addressing these goals is for students to be able to learn content in such a way that they can apply it in a variety of situations (promote transfer) and to become skilled science learners, becoming competent at participating in the practices of science and science learning.

Our assessment goals have been (i) to provide alternative, more dynamic ways to assess student learning and skill mastery beyond static measures like standardized achievement tests; (ii) involve our teachers as collaborators in each step of this developmental process; and (iii) validate those processes and tools that would have value to other teachers of middle school students.   A lesson we continue to acknowledge as we work with our teachers is that we all want students who have high achievement tests scores, but we also want students who are learning the deep principles and processes of science that will go on to become the innovators, designers and problem solvers of tomorrow. We want to extend what counts as learning in science to include the process skills that will equip our students to understand science, not just learn facts.

Indeed, the science education community and the recently-published American standards about science literacy want students to gain both of these competencies (American Association for the Advancement of Science (AAAS), 1993) – to learn science concepts in a way that allows them to apply those concepts to new situations as they arise and to become enculturated into the practices of scientists. This includes being able to enact science process skills such as inquiry, observation, measurement, experiment design, modeling, informed decision making, and communication of results, as well as social practices such as learning from each other. The aim is for students to learn in ways that will allow them to take part in those practices in skilled ways both inside and outside the classroom.

Transfer, in the context of science learning, means gaining expertise in the practices of the scientific community (e.g., designing experiments, managing variables, justifying with

evidence, analyzing results, planning investigations, communicating ideas, communicating results, incrementally building up one's understanding of a new concept). Transfer also means learning scientific concepts and their conditions of applicability in order to engage in scientific reasoning.

Assessment and Scaffolding

We have worked in collaboration with our teachers to develop a series of assessment procedures to capture not only the science content knowledge that our students acquire, but the process skills so important to metacognitive skill development, problem solving, collaboration, and research skill development. Our *design diaries* (Puntambekar & Kolodner, in preparation) support and scaffold the individual student throughout the design process. They also provide a window into the cognitive activities of the student for the teacher to use as an alternative assessment tool. We have also developed a *self-assessment survey* for students that was refined with our teachers and taken from a larger data base set of potential assessment categories and skills that are available to the individual teacher's need and to the growing repertoire of skills of her students. A third assessment tool is more dynamic in nature and involves defining and refining a series of mini-problems for students to address over one or two class periods. These problems (*performance assessment tasks*) were adapted from the PALS (http://pals.sri.com/). We've adapted a series of such problems that can be administered across the school year to plot developmental changes in students abilities to "do" science.

From several iterations on the development of our assessment and scaffolding tools, and from the data we've collected to show student learning with these tools, we have had an impact on addressing the national standards for middle school students in science, as listed below.

- Design an investigation
- Communicate investigative process
- Answer original question
- Communicate results
- Understand relationship between explanation and evidence
- Use tools to gather, interpret and analyze data (including mathematics and computers)
- Understand explanation, in contrast with mere description
- Consider alternative explanations presented to them (not necessarily ones they themselves come up with)

The assessment and scaffolding tools we've designed address these standards and more. Our results so far show that our *Design Diaries*, our *Student Success Handbook* (Gray, Groves, & Kolodner, 2000) and our assessment instruments have potential for impacting the development of these very important skills, routinely practiced in our LBD™ classrooms. We also note from the standards (AAAS, 1993) several practices that students tend to have difficulty with, and LBD provides extensive practice opportunities and scaffolding for those:

- Identifying variables
- Controlling for more than one variable
- Understanding that different variables have different levels of effect (lots, little, none)
- Mentioning evidence against their conclusions
- Having conclusions that differ from previous opinion
- Coming up with alternative explanations.

LBD units address specific science content as well, and our units actually cover more content and process skill development than these standards, for example, our students develop great skill in collaboration. The list does, at least, highlight the kinds of things we're reciprocally assessing and scaffolding. Our assessment and scaffolding tools are developed to assess and support the acquisition of these skills and "habits of mind".

Performance Assessments

Scientific reasoning skills are difficult to measure, but we have had some success in showing the acquisition of these very skills by the students in our LBD™ classrooms. As others have recognized, general strategies emerge out of repeated experience in problems rich in domain content. To fully understand the emergence of scientific reasoning skill, prior knowledge will need to be included in the picture (Zimmerman, 2000). The challenge will be to continue to explore the role of prior knowledge in the acquisition of new knowledge. We have developed a series of problem solving tasks for students in our work to assess their scientific reasoning as well as their collaboration skill. We have gained a great amount of expertise in the past two years in designing performance tasks that can be used to assess student learning of skills and practices and in creating rubrics that can be used to analyze the extent to which students are participating. Preliminary evidence based on these tasks shows that LBD™'s practices do indeed promote transfer

in the subset of the students we have evaluated, showing us that such tasks can be used for assessment of skills learning and can be coded reliably.

We have adapted performance assessment tasks to follow a format that allows us to better assess the collaboration and science process skills that we seek to promote in the LBD™ curricula. The task is designed in three parts: (I) students *design an experiment* to gather evidence to address an issue in the context of a real-world problem; (ii) students work in groups to run a specified experiment with materials we have provided, and gather data from this experiment; (iii) students answer questions that require them to utilize the data they gathered, and to apply their knowledge of science to interpret the data. The quasi-experimental design has different classes assigned to different participation conditions: Some classes have the students do all three parts of the task as a group, writing a single group answer; some classes have the students run the experiment as a group, but to work as individuals on parts 1 (designing/writing an experiment) and 3 (interpreting data, answering questions); and some classes have the students work together on all three parts to develop answers, but each student writes these answers in his/her own words.

We videotape the two conditions in which groups of students work together throughout the task. The design-an-experiment part of the task allows us opportunity to judge group ability to design an investigation, their understanding of what a variable is, and their ability to control variables, among other things. The middle part helps us determine their ability to carry out a procedure carefully and correctly: to measure, observe, and record. The third part allows us to determine if they know how to use evidence to justify and how well they can explain. All three parts provide evidence about their collaboration and communication capabilities and their facility at remembering and applying important classroom lessons.

An example task may help bring this to life. In "Where the Rubber Meets the Road," (http://pals.sri.com/),   part I has students design an experiment that compares the efficacy of two tire types that differ in the hardness of the rubber used when tested in different road conditions. The science concept being tested is understanding of the force needed to overcome sliding friction.

Coding categories include negotiations during collaboration; distribution of the task; use of prior knowledge; adequacy of prior knowledge mentioned; science talk; science practice; and self checks during the design of the experiment, and each group is scored on a likert scale of 1 - 5, with 5 being the highest score. (See Appendix 1 for the coding scheme developed to assess collaboration and science practice skills during these tasks.).

Preliminary results

When we used our coding scheme to analyze student performance, we found that for an LBD typical achievement levels classroom vs. a similar comparison classroom, there were statistically significant differences in mean scores for the "distributed" and "self checks" measures, and a nonsignificant trend for prior knowledge adequacy.[1] In each case, the LBD means were higher than the comparsion class. For the LBD advanced-achievement classroom vs. a similar comparison classroom, there were significant differences for the negotiation, science practice, and self-check measures with higher LBD means (See Table 1). LBD students were better than comparison students at collaboration, metacognitive awareness of their practices, and ability to remember and use what they had learned previously. Students in LBD classrooms participated in collaboration that can be characterized by negotiation and the distribution of the work. Students in comparison classrooms worked in groups without taking advantage of the unique affordances when work is distributed or solutions negotiated..

This assessment is important for several reasons. First, it tells us that the combination of scaffolding and orchestration that we have developed for LBD is successful in achieving our student learning goals. This adds to less formal evidence presented above that design diaries are providing the kinds of scaffolding we predicted. Second, it tells us that we are on the right track in designing the performance tasks and their coding metrics. As these become more concise, we will make them available to teachers and students as scaffolding, showing them the kinds of articulations and practices we expect them to be able to achieve. And third, it provides evidence that these "habits of mind" are being learned and transferred (Kolodner, Gray, & Fasse, submitted).

---

[1] Reliability for the coding scheme ranged from 82-100 percent agreement when two coders independently rated the tapes. For this set of data, a random sample of four tapes for three teachers and three tapes for one teacher (complete set) were used for this analysis. A total of 15 group sessions were coded representing 60 students.

References

American Association for the Advancement of Science (AAAS). (1993). Benchmarks for science literacy. Project 2061: Science of all Americans. Washington,DC:Author.

Gray, J., Groves, H. and Kolodner . J. L. (2000). A survival guide: The Student Success Handbook for learners in project based science environments. *Proceedings of the International Conference of the Learning Sciences* (ICLS 2000), Ann Arbor.

Kolodner, J. L., Gray, J., and Fasse, B. (in preparation). *Promoting transfer: Rituals and practices in Learning by Design™ classrooms*. Georgia Institute of Technology, College of Computing, Atlanta, GA.

Kolodner, J. L., Crismond, D., Fasse, B. B., Gray, J. T., Holbrook, J., Puntambekar, S., & Ryan, M. (in press). Problem-Based Learning Meets Case-Based Reasoning in the Middle-School Science Classroom: Putting Learning-by-Design™ into Practice. *Journal of the Learning Sciences*.

National Research Council (1996). *National science education standards*. Washington DC:

Puntambekar, S., Kolodner, J.L., (in preparation). From 'Learning to Scaffold' to 'Scaffolding to Learn': A Case for Distributed Scaffolding.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20. 99-149.

Table 1:

Means and standard deviations for categories from performance assessment coding for LBD students (typical and honors) and Comparison students (typical and honors)

Coding category	Means (SD) LBD Typical	Means (SD) Comparison Typical	Means (SD) LBD Honors	Means (SD) Honors Comparison
Negotiations	2.50 (1.00)	1.50 (.58)	4.50 (.58) ***	2.67 (.58)
Distributed Effort/tasks	3.25 (.50) *	2.25 (.50)	4.00 (1.15)	3.00 (1.00)
Prior knowledge	2.25 (.50)	`1.75 (.50)	3.75 (1.50)	3.0 (.00)
Prior Knowledge adequate	2.75 (.96)	1.50 (.58)	3.50 (1.00)	2.67 (1.15)
Science terms used	2.50 (1.29)	1.75 (.50)	3.50 (1.00)	2.67 (1.15)
Science practice skills	2.75 (.96)	2.25 (.50)	4.75 (.50) ***	2.67 (.71)
Self-checks	3.00 (.82) **	1.50 (.58)	4.25 (.50) ***	2.33 (.58)

Significance levels: * = *p* < .03; ** = *p* < .02; *** = *p* < .01

The means are based on the likert scale: 1 - 5

Appendix 1:

Performance Assessment tasks: Coding for science practice
Jackie Gray, Paul Camp, Jennifer Holbrook, Barbara Fasse, and Janet Kolodner

Additional notes are fine and can be recorded on the coding sheet.
Please note which event segment is being coded for each episode:
planning an experiment; problem set up; experimental manipulation; response to written questions.

In general, the 5 -point likert scale reflects the following quantitative continuum. Details for each item are also included below.

1 = Not at all: no evidence of the quality to be rated

2 = Some evidence that at least one episode or one student exhibits the quality rated

3 = The quality is exhibited half the time

4 = The quality is exhibited for more than half the episodes

5 = The quality completely captures the nature of the episodes

Design an experiment segment:

Within an episode, the context of the group is characterized by:

Negotiations

Not at all	at least one of the members of the group suggests a compromise about some aspect of the procedure	at least one of the members of the group suggests that compromise or debate is needed for at least half the issues that require it	at least two of the members of the group questions several aspect of the procedure and the group makes the needed change	Most decisions are made about procedure by the entire team contributing and decision making is consensual
1	2	3	4	5

Distributed efforts and tasks

Not at all	at least one of the members of the group suggests that others help do the task	at least two of the members of the group suggest that all do some part of the task	at least one of the members of the group suggests and leads the group in dividing and doing the task	More than one of the members of the group enlists the participation of all the team in doing the task
1	2	3	4	5

Level of Understanding of the problem

Not at all	The group thinks the task is to write something down disregarding the "design" aspect	at least two of the members of the group try to work out a method and "run an experiment" with the material available	at least one of the members of the group recognizes that an experiment is to be designed and shares with the other members	More than one of the members of the group enlists the participation of all the team in designing an experiment and that it calls for additional materials
1	2	3	4	5

Use of materials to get to a method

Not at all	At least one member of the group manipulates the material (s) while trying to develop a solution	at least two of the members of the group examine and use the material in a way that might suggest an effort to discover a method	at least two of the team members manipulates the material to explicitly suggest a method	The team explores the material as if messing about to understand what to include in their design/method
1	2	3	4	5

Prior knowledge is defined as students referring to some aspect of the curriculum unit that relates to the current problem; referring to some aspect of a personal experience that seems to relate to the current problem; referring to some aspect of the science concept or method at issue that appears to come from previous exposure to the concept or skill.

Students show evidence of using prior knowledge to solve the problem

Not at all	at least one of the members of the group mentions a prior event or concept that relates to the problem	at least half of the team mentions a prior event or concept that relates to the problem	Several events and concepts are mentioned and applied to the problem	The group routinely recalls events or concepts that assist in their collaborative problem solving
1	2	3	4	5

Prior knowledge seems adequate

Not at all	at least one of the mentions of prior knowledge is followed up on and is useful	At least half the mentions of prior knowledge are appropriate to the problem	More than one member of the group mentions or follows up on events or concepts that are useful	Every mention of prior knowledge is directly applicable to the problem
1	2	3	4	5

Science terms are used in a way that indicates some degree of understanding and can be argued that they are not attributed to the science terms included in the problem description.

Students use science terms to discuss problem solution

Not at all	at least one of the members of the relates the discussion to some science concept	at least half the team relates the discussion to some science concept	Most of the team members use science concepts or terms in such a way that accurate understanding and application are noted	All members of the the team members use science concepts or terms in such a way that accurate understanding and application are noted
1	2	3	4	5

Students use science practice to decide on method/procedures

Not at all	at least one of the members of the group suggest a method to test at least one variable	at least one of the members suggest a method and indicates an understanding of fair testing	at least one of the members suggest a method and indicates an understanding of fair testing and controlling for variables	Most of the team agrees that the method used will fairly test the important variables and their decisions would actually be a reasonable experiment
1	2	3	4	5

The episodes are characterized by group self-checks on procedures

Not at all	at least one of the members of the group questions some aspect of the procedure	at least one of the members of the group questions some aspect of the procedure and the makes the needed change	at least one of the members of the group questions several aspect of the procedure and the group makes the needed change	More than one of the members of the group questions several aspect of the procedure and the group makes the needed change
1	2	3	4	5