WRITING FOR SCIENCE, SCIENCE FOR WRITING

A STUDY OF THE SEATTLE ELEMENTARY SCIENCE EXPOSITORY WRITING AND SCIENCE NOTEBOOKS PROGRAM

Laura Stokes
Mark St. John
Jo Fyfe

With Assistance From:

Kathleen Dickey
Barbara Heenan
Judy Hirabayashi
Dawn Huntwork
Dianne Maxon
Pamela Tambe

Inverness Research Associates
July 2002

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Expository Writing and Science Notebooks Program is of interest because of its deliberate approach to improving the use of science notebooks for the twin goals of improving learning in science and developing expository writing skills. The program consists both of a curriculum for writing within the context of the science kit program, and a strong supportive professional development program.

For this study, we examined the extent to which teachers value the program's curriculum and teaching strategies, and the extent to which they believe it benefits their students. We also carried out a direct study of the science notebooks of a sample of teachers who are implementing the writing curriculum in their classrooms.

Two groups of experts rated the work in the student notebooks. A group of Lead Science Writing Teachers in Seattle assessed the extent to which the work in notebooks reflects the goals of the program. A panel of independent experts made judgments about the nature and quality of student learning opportunity from the perspective of the larger national reform community.

Our study addressed three broad questions:

1. To what extent and in what ways does the program assist teachers in improving their teaching of science and writing?

2. What evidence is there that the program is adding real value to students' opportunity to learn science and writing?

3. How can the program continue to refine and develop itself so that its quality is improved and its reach is extended?

Summary Assessment

- The Expository Writing and Science Notebooks Program is meeting the practical instructional needs of teachers using science kits. The program thus adds considerable value to, and is an important dimension of, the kit-based science program recently implemented in Seattle schools.

- The program is contributing to student learning of science and writing across the elementary grades in ways that are consistent with the program's goals and standards, and that are also valued by the broader field.

- The program appears to be especially effective in supporting the learning of students with special needs and fostering development of students' emergent literacy. The program thus seems to have potential to contribute to more equitable outcomes for different student populations.

- The Expository Writing and Science Notebooks Program has potential for large-scale implementation because it has a well-specified curriculum, systematic professional development, and growing teacher leadership capacity.

- From the perspective of some Seattle teachers and all members of the independent panel, the program currently appears to under-emphasize uses of writing that support students' own inquiries and their pursuit of their own scientific thinking.

Implications for Further Development

We offer the following suggestions for the next phase of the program:

- Continue offering teachers across the district access to the program materials and its approach (centrally and / or at school sites) in order to grow the number of teachers involved in the program.

- The supplemental writing curriculum is currently designed so that writing tasks are intended to be optimally supportive of key science ideas in each kit, and this should continue. Further refinement and expansion of the writing curriculum should aim at an additional purpose: offering students greater opportunity and challenge in terms of pursuing their own inquiries and scientific thinking.

- Offer participating teachers ongoing opportunities for professional development beyond the initial set of workshops. These opportunities should include joint study of student work in notebooks and shared problem-solving about classroom strategies, including adapting the writing curriculum for students with different skills.

- The program can and should offer accomplished teachers opportunities and roles for leadership that go beyond those that now exist. Key Lead Science Writing Teachers should, for example, play a major role in expanding the writing curriculum to foster student-centered inquiry, as well as facilitating ongoing professional development for participating teachers.

We believe it would make sense to discuss the possibility of continued evaluation by Inverness Research Associates after the program's next cycle of work is fixed and after everyone involved has had an opportunity to reflect on the results of this study.

WRITING FOR SCIENCE, SCIENCE FOR WRITING

A STUDY OF THE SEATTLE ELEMENTARY SCIENCE EXPOSITORY WRITING AND SCIENCE NOTEBOOKS PROGRAM

I. BACKGROUND

The Expository Writing and Science Notebook Program as a Model of Interest

There is considerable interest both in the Seattle district and the national reform community in improved teaching and learning in science, in literacy development per se, and in literacy development within disciplines. In our studies of many elementary science programs, we have observed that science notebooks are widely considered to be an increasingly important component of a hands-on science program; however, we have seen no program that has made the kind of concerted effort that the Seattle district has to support teachers' consistent and purposeful use of notebooks.

This program is of interest because of its deliberate approach to improving the use of science notebooks for the twin goals of improving learning in science and developing expository writing skills. The program is making a clear attempt to achieve a symbiotic relationship between science and writing, with the teaching and learning of each reinforcing the other. We believe lessons from the program can inform the program itself, the funder, and also larger national efforts to improve both science learning and literacy.

The Expository Writing and Science Notebook Program has three components:

- Supplemental curriculum for expository writing specific to each of the 18 hands-on units. These included focus questions for writing, as well as thinking / writing frames and graphic organizers designed for the specific lessons in each unit.

- Professional development that is available to large numbers of teachers across the district. This component consists of series of four workshops per grade level band that introduce the overall approach and apply it to each specific science unit / kit.

- Teacher leadership development for three to five teachers per grade level. Lead Science Writing Teachers (LSWTs) assist in developing and field-testing curriculum strands and materials for workshops.

We at Inverness Research Associates have studied the program for two years. During the first year we familiarized ourselves with the program by observing professional development workshops and interviewing key staff members; we also conducted in-depth interviews with a sample of twelve teachers identified as having varying degrees of participation in the program. In the second year, we focused our study primarily on the student notebooks of students whose teachers are participants in the program. These notebooks, along with teacher interviews, are our key sources of data for this report.

The Evaluation Study

Our study to date has been an effort to begin assessing the promise and potential of Expository Writing and Science Notebook Program as a contributor to teaching and learning in Seattle's elementary science program.

We focused on the following broad questions:

- To what extent and in what ways does the program assist teachers in improving their teaching of science and writing?

- What evidence is there that the program is adding real value to students' opportunity to learn science and writing?

- How can the program continue to refine and develop itself so that its quality is improved and its reach is extended?

We have taken a two-step approach to the study. Our first step was to examine the extent to which teachers see the need for, and the value of, the new supplemental curriculum and teaching strategies related to the teaching of writing in the context of science, as well as the extent to which they can actually put into practice what the professional development offers them. In-depth interviews with a range of teachers gave us insight into these questions.

The second step was to examine the notebooks of teachers who both participate in the program and are implementing the writing curriculum in their science instruction. The science notebooks of students are work products that reflect a very wide range of influences—some related and some not related to classroom instruction, and some related and some not related to this specific professional development program. Nonetheless, we believed it would be possible to see evidence and patterns in the notebooks that reveal what we could identify as teacher implementation of the quite explicit and structured curriculum and teaching strategies of the program. We found that we in fact were able to observe clear influences of the program on what students

were writing about and how they were writing about it, and could see how their writing was interacting with their learning and thinking in science.

We designed the study so that two expert perspectives would be brought to analysis of the student work. First, we wanted to assess the degree to which the work in student notebooks reflected the program's own goals and standards. To do this, we invited lead teachers from Seattle to rate the work in notebooks using the program's own criteria and standards. Second, we wanted to assess the degree to which the student work, and the program's goals and standards, are consistent with the vision for science teaching and learning and for literacy held by experts in the field. To do this, we invited a panel of independent experts to study the student notebooks. The panel included practicing elementary teachers of science and of writing in the context of science, professional development leaders and reform activists in elementary science education, and researchers of both science education and writing. We thus used student notebooks as a "window" onto the program, relying on experts both inside and outside the program to make judgments about the nature and quality of student learning opportunity reflected in the student work.

This report

The report is intended to provide feedback both to the funder and to project leaders about the opportunities the program is providing to teachers and to students. It is also intended to assess the program's potential to continue serving elementary science education in Seattle.

Three major sections of the report follow this introduction. Section II presents teachers' perspectives on the program—its value to them and to students, and the opportunities and challenges they face in implementing it. Section III presents findings from the study of student notebooks. (In that section we detail the design of that aspect of the study, including the sample of participating schools and the scoring criteria.) Section IV summarizes the major findings of the study and offers our recommendations for the next phase of the project.

The Appendices include the teacher interview protocol, the scoring criteria and rubric used for the notebook rating sessions, samples of student work from notebooks to illustrate the criteria, and results of the teacher background survey.

A caution

There are many factors that affect student work in notebooks and that thus confound analysis: the teachers' foundational teaching knowledge and skill, the students' foundational knowledge and skill, the structure and content of the science kits / units to which the writing curriculum is linked and which it serves, the constraints of school-site

schedule and priority that limit attention to the teaching of science, the distribution of resources to students with special needs, including programs that pull them out of classrooms—the list goes on and on. As one of the independent raters said, "The notebooks are a product of the prompt, the lesson, the kit, the kid, the instruction, there are a lot of things that shape what shows up in the notebook."

This study, because it is relatively small in scale and because it cannot control for or eliminate other determinants of the quality of work in notebooks, offers evidence that is suggestive of patterns and influences rather than hard proof.

II. TEACHERS' PERSPECTIVES ON THE EXPOSITORY WRITING AND SCIENCE NOTEBOOKS PROGRAM

In early 2001, we visited 12 teachers at their schools and conducted in-depth interviews with them. The interviews focused on their approaches to teaching writing in science, the value of the professional development they received, their ability to put into practice what they learned, and the benefits they believe the program offers to their students.[1] In May of 2002, we interviewed 11 Lead Science Writing Teachers after they had read student notebooks from grades 1, 3, and 5. We asked them to share insights they had gained from reading the student work. We also asked for their reflections—both as classroom teachers and as leaders in the program—about the role of this project as a contributor to their teaching and to Seattle's reform effort in elementary science. They shared with us their beliefs about the program's value, and also their ideas and hopes for its future.

This section summarizes what we learned from these teachers.

A. Contributions to Classroom Practice

We wanted to find out whether teachers value the program and find it useful in serving their practical needs. Their responses were overwhelmingly positive; participating teachers attribute much of their confidence and ability to teach science—and their enjoyment of teaching it—to the Expository Writing and Science Notebooks Program.

Overall finding: **Teachers state strongly that the Expository Writing and Science Notebooks Program gives them the resources and strategies they need in order to improve their teaching of the hands-on science kits adopted by the elementary science program, and to improve their teaching of writing. The approach also helps them assess student learning and monitor their own teaching.**

The teachers we interviewed stated repeatedly that the writing curriculum and workshops give them concrete and specific strategies that they can use, and that will help them teach the kits. This help was especially important because many of them lacked experience and confidence in teaching hands-on science:

> When I started teaching the kits, it was overwhelming—I couldn't figure out how to get them to put quality things in a notebook. Then I had the opportunity to go to some of Betsy's writing classes and get the basic frames and the basic idea, and the examples of

[1] The protocol is included in Appendix I.

> how it had worked for other kids. Then I felt more like,' okay, I can do this.' [4th grade teacher]
>
> I do not have a science background. It hasn't been something that I had loved and enjoyed as a child and through college. So for me this program made it possible for me to become a teacher who loved and saw the value in teaching science because I was so supported in it. If I had been left to my own devices, I don't really know what I would have done. [5th grade teacher]

The teachers say that the writing packets are the most valuable resource they have for effective teaching of the kits:

> A lot of teachers who have been through the expository writing classes read the writing packets more than the kit manuals now. The writing packet has all of the helpful hints that help that lesson go successfully–what works and doesn't work. [5th grade teacher]
>
> I don't have a hard science background and I don't know the kit stuff really well, I don't have a lot of the adult content internalized, so I can't pop up with things that I would like them to write about. Having the frames helps me think about what I am really asking them, that essential question for the understanding. It focuses me as much as it focuses the students, I think. [5th grade teacher]

The teachers find that the approach to writing improves their ability to assess student learning:

> The writing helps in my assessment of kids' learning and whether they are really grasping the concepts, because they have to express themselves and answer the focus questions, which gives you an idea of what they have really understood in the lesson. You can have a class discussion, but sometimes it is hard to monitor and figure out who in the class has that conceptual understanding. [3rd grade teacher]
>
> When they use the frame, you can actually tell better what it is that they are understanding, or what they are confused about. [4th grade teacher]

Some of the teachers also said that the new teaching tools and behaviors they learned from the program helped them develop a more deliberate and more student-centered approach:

> My teaching is more intentional through things I have learned in the writing program. I began to really analyze my teaching, because the way I taught before, I would talk to the kids too much and taught them things that were very abstract. But I am restructuring my teaching totally so the kids are doing a lot more of the work. That has come through the writing and science, and just seeing how important it is for the kids to set up their investigations and take ownership of what they are learning, to have a personal influence in what they are doing. [4th grade teacher]

One teacher said that the notebooks help her provide evidence of teaching and learning to parents:

> I have a father who is an entomologist, he studies bugs. When we did the Organisms unit and I sent home the science notebook, he came back with it and said, 'I cannot believe my child is looking at the bugs and writing about it.' He really felt invested. [1st grade teacher]

B. Contribution to District Reform

Teachers often feel that their districts make demands on them and hold out high standards, but do not offer supports that help teachers reach them; we know that many teachers in Seattle share this view on the whole. We asked teachers for their perspectives on this program with respect to district expectations.

Overall finding: **Teachers state that the expository writing program is necessary if they are to respond to district expectations for improving both science and literacy. They say the offerings in science stand out as a truly valuable district-sponsored professional development program. Further, they believe the science notebooks reflect the kind of rigor the district would like to foster in teaching and learning**

A number of the teachers we spoke with said they appreciate the fact that the district offers high quality professional development in science that helps them improve their practice with respect to standards.

> The science is heads and tails above anything that goes on in any other kinds of professional development in the district. [3rd grade teacher]

> I think it is a really complex thing that they are asking us to do, to put science and writing together so students come to some kind of real conceptual understanding. They are not only asking the kids to learn how to do it, but they are asking us to learn how to teach it when we really haven't got a strong background in it, that is why the professional development is just so crucial, and has to be ongoing and it has to be deep. There is hardly time or money for either. [5th grade teacher]

The teachers also believe the approach to science notebooks increases the rigor of the science program to the extent that it exemplifies district expectations:

> The district's whole thing right now is academic rigor, it's this big new buzz word and I don't think anybody knows what it means. I think this is the program that shows the most academic rigor. If they want to know what rigor looks like, they need to look at these notebooks. [5th grade teacher]

Teachers understand that the writing program is adjunct to and serves the science program. They feel strongly that professional development in both the kits and the writing are necessary and mutually supportive:

> The science resource teachers that teach the initial-use workshops, I think they are a critical part of the success of it. There is a real fear about the commitment of the district to have quality initial-use workshops, but we really need both. At the initial use workshops you get the concepts for the kits, but we can't get into the depth about specific lessons—but that is a place where any teacher says, 'well how would you have them write about it?—and we can say, 'you can go to the writing workshops.' The classroom teachers have their plates full and they can't come up with everything on their own but if you can give them the tools, they can use it. But the science workshops need to be there too because the science, that is what engages the kids with the writing. [5th grade teacher]

C. Benefits to Students

For all of the teachers who participated in this study, the program's potential to improve student learning was what mattered most. They eagerly spoke to us about the ways they believe their students benefit.

Overall finding: **Teachers believe strongly that the program's approach to writing in science notebooks helps their students learn both concepts and skills in science, and learn to write in ways that deeply engage students and that also reflect the rigor of science as a discipline. Teachers believe the approach is especially powerful for English Language Learners and others for whom writing is a struggle. Teachers also believe that the skills build cumulatively, are lasting, and support students' learning across the curriculum.**

Teachers observe that the writing supports students' formation of science concepts:

> To write about the concepts they have to think about them. That is the whole idea–that the writing has to come from their thinking. If they didn't have to actually write out how you can make a long string have a high pitch, they may never have really thought about that. It forces them to think about the concepts in science. [3rd grade teacher]

The teachers believe that the structured approach to the notebooks introduces students to the authentic skills and thinking processes of science as a discipline:

> I think the first thing is for students to understand that scientists keep meticulous records of what they do, and if they don't, they are not really doing science. They can't come to

> some kind of conclusion, present it at a meeting or present it to a colleague if they do not have some kind of evidence to back up what those conclusions are. This is one way to give students an idea of what that feels like. The other thing is to help them learn to process their thinking. If you have to write the process that you went through while you were thinking through an activity or an experiment, and put into words what your conclusions are, it helps you process it. [5th grade teacher]

Teachers say the program's approach offers students specific skills for thinking and writing, and with a rigor that reflects science as a discipline:

> It gives them another tool for expression that before this point, they would not have. They don't know how to explain their reasoning, for instance, unless we teach them. They don't know how to analyze data and interpret it, unless we teach them. This is what the writing is doing for kids. [3rd grade teacher]

> It gives them an idea of how to organize what they are thinking [4th grade teacher]

> I have seen other writing approaches where the assignment would be something like 'pretend you are a sow bug in your terrarium, describe your day,' or something like that. That is so different than what we are asking, that's more of a fictional write. What we are asking them to do is really deep, and it is hard to learn to do it well. Pretending you are a sow bug in a terrarium would be easier to deal with than to figure out, you know, 'I notice that the sow bugs are doing *this* and my evidence is *that*, and therefore I think *this*.' And that is my goal, to get them being able to do that, to see what evidence really means. If my student says 'I believe that the sow bug is half-shedded,' well what is her evidence? And she can write it in there–'because half of it is cloudy white and half of it is black.' That is the kind of thing I am going for. [5th grade teacher]

With respect to development of writing skills *per se*, teachers believe the writing and science notebooks program offers them a valuable alternative to a writing curriculum that asks students to draw primarily from their imagination or personal memories. The concrete phenomena and immediate experiences of hands-on science give students meaningful content to write about:

> I like teaching the writing in conjunction with the science curriculum because it makes the writing meaningful for the children. In my first year, I didn't have all of this training in the writing and I really didn't know how to go about teaching the writing. I had them write in their journals and write stories, but they didn't get into it as much because it wasn't meaningful. Playing with these balls and bouncing them was something they were experiencing firsthand. They have just done the bouncing, so they can go back and read what they observed, and write about it right away. [1st grade teacher]

> A lot of writing kids do in school is, I hate to say it, meaningless to kids, where they just read about something and regurgitate it. But this is real–this is something they are doing, something they are touching, this is something that they are understanding from their own observations. [3rd grade teacher]

> For some kids writing from something concrete in front of them is a more powerful place than to tell them 'use your imagination.' Certainly there are children who are wonderful

> at creating amazing stories, but I have found that kids who really struggle with writing all of a sudden have the ability to do more, because they just had this experience that happened five or ten minutes ago, and they can connect with that, right there. [5th grade teacher]

Teachers say that what students learn from writing in their science notebooks supports their skill development and inclination toward evidence-based reasoning in other subjects:

> It is helping them with their expository writing across the board, because in social studies, math, other areas of the curriculum, they are getting more practice in writing expository. You see them using transition words and the frame structures in other subject areas, it carries over. They use more deductive reasoning, too, like for example, in social studies we do a 'mystery country' activity where kids get certain clues, and when the kids are in the discussion, they will automatically say, 'well I think it is Japan *because* this, and this, and this. They provide the evidence, which we force them to do in a science notebook. I think the reasoning just naturally carries over into the other content areas. [3rd grade teacher]

> It covers a lot of math skills. I haven't had to do measurement, because they are weighing balls, they are measuring their plants. They are getting a lot of skills that you don't have to teach later. [1st grade teacher]

Third grade teachers, who feel strong pressure to prepare their students for the WASL, believe the program is helpful to their students:

> I don't know if you have seen examples of our WASL, the 4th grade Washington Assessment of Student Learning, but most of the math sections, there are a couple of multiple choice, but for the most part, they are short answer with extended response, and the students have to explain their thinking. This gives them the ability to do that. [3rd grade teacher]

The teachers have especially strong beliefs about the power of the approach to support the learning of students who, for a wide range of reasons, find it a struggle to use written language. They may be first-graders just learning to write, or English Language Learners, or students with learning disabilities, or students with such under-developed motor skills that they resist writing. The teachers believe these students develop the skills and confidence they need to participate and communicate alongside their peers:

> It really helps those kids who just don't know where to start, maybe they are ESL or Special Ed. Some of my more advanced students, if I ask a question, will automatically know how get started, but the kids who don't will just sit there. If they didn't have that frame or if they didn't have the modeling, they wouldn't be very successful at all. It gives kids who don't have the ability to write, that voice. [3rd grade teacher]

> I think it gives kids power, it opens up possibilities for them, they can express themselves. You know that life is getting better for Special Ed or ESL kids when they

> come to you with their science notebook to say, 'look at what I just wrote.' [5th grade teacher]
>
> Within the first few weeks of me teaching my first science unit, I was able to go to one of Betsy's classes and she started introducing the frames. I thought, 'well I don't know what I feel about frames, because aren't they supposed to be thinking independently?' But I was overwhelmed and new to the science curriculum, and I had a high level of bilingual kids, so I began to use the frames. What was amazing to me is that the my students felt so much more comfortable expressing themselves that way. I even pulled it over into math, and for my students, their math writing shot way up, from struggling to get them to write a sentence, they were suddenly able to write an entire process of how they solved a problem [4th grade teacher]
>
> For the ones that who aren't good at reading or writing, this is engaging them and making them excited. Then they find out 'I can do this!' and then there is a snowball effect and they keep trying. I've noticed kids lately, when they have some extra time, they will go back and read their science notebooks, because they think it is fun. It is powerful. [1st grade teacher]
>
> You can tell some of these students with the terrible handwriting, they lack coordination and motor skills. They usually write the smallest they can get away with, they are just resistant. But I was astonished at the amount of writing in some of their notebooks. [3rd grade teacher]

Program participants observe cumulative development of skills with notebook writing, including students' self-initiated use of notebooks and their ability to wean themselves away from the thinking and writing frames to which they are initially introduced:

> I'm in a high poverty school. I have 20 children,14 of them speak another language at home, and 9 of them go out for bilingual instruction services during the day. Last Friday I saw an amazing thing. I had them exploring with the balls and bouncing them. By now [5 months into the school year], they know that when they start science, they get out their notebooks, they write the date at the top, and always write a title. So they had their books out and they did free explorations with the balls, and I didn't tell them to write in their notebooks, we hadn't talked about it. But about half the kids just started on their own writing about what they are finding out about their balls. I hadn't told them to do that, so they are just starting to write observations on their own. I found that amazing. [1st grade teacher]
>
> There is a bit of concern with a few teachers in my building about the writing frames being a crutch for kids. I can definitely say they are <u>not</u> a crutch. It gives kids a frame to start and all of the students benefit from it, from the advanced to the struggling ESL or learning disabled kids, but the kids who need to continue using it do, so they at least get something on paper and their writing is organized, their thoughts are organized. And the students who don't need to continue it instantly jump beyond the frame and their writing is that much better. [3rd grade teacher]
>
> Once they have gotten more of a handle on it, they tend to do more on their own, or they add more in or they completely leave the frame. But the nice thing is, once they have had the frames, they tend to be more organized later on. [4th grade teacher]

D. Ideas and Hopes for the Future

These participating teachers are heavily committed to and invested in the program. They have both hopes and concerns about its future, as well as concrete ideas about next steps they would like it to take.

Overall finding: **Teachers believe the program has potential to have a broad impact on student learning over time. However, they are concerned that competing demands on instructional time can impede in-depth implementation of both hands-on science and the writing-in-science program that adds so much value to the science kits. In addition, they believe the program needs to increase its attention to supporting student-generated inquiry and to offer new and additional professional development support.**

Program participants are confident that the program has potential to have a broad and positive affect on teaching and learning. One teacher said:

> I think it takes a number of years before you see a change across the board. You can introduce a program and hope that you will see amazing results in one or two years, but in reality, it is going to take a long time. But I bet if you had the ability to blindly pick samples from 6 or 7 years ago and compare them to these, I am sure you would notice some improvement. I am pretty confident that for all of my students, notebooks have improved dramatically. [3rd grade teacher]

Because the writing program has a clearly laid out approach and supplemental curriculum, the lead teachers feel well-supported as they work to spread implementation.

> I am in charge of science in 1st grade, and so I can go through the writing packets with the 1st grade teachers, and say, okay, this is what we need to get out of this unit–this is the kind of writing we are looking for. It creates enthusiasm for the unit. [1st grade teacher]

The teachers point out, though, that the myriad pressures on their teaching time will always play a strong role in supporting or obstructing in-depth implementation:

> The biggest concern–you will hear it across the board throughout the district from every science teacher–is the time. And especially if you really work on the writing, it is very time-consuming. A 45-minute lesson all of a sudden turns into three or four 45-minute lessons. And that is a huge chunk of time, considering we are supposed to teach three units and each unit has 16 lessons in it. Unfortunately, it is not like you can say 'we can skip math this week,' because we have standards in those areas, too, that we have to meet. [3rd grade teacher]

> The time is a challenge because I value this process of them making their own charts and writing out their own explanations, but it does take more time than just giving them a handout and saying fill in the blanks. So I have to make sure that I am thinking about that. Even though I am doing science every day, I can't finish the unit in a couple of weeks because I want to leave time for the students to work in their notebooks. [4th grade teacher]

Because Seattle uses school-based decision-making, school-site priorities have a very strong influence on teachers' use of instructional time. The teachers' comments below show how this is an obstacle to consistent implementation:

> What's great about my school is everybody does science, so that there is a commitment to that. The way that our time is structured, you can have hour blocks or hour and twenty minute blocks, which makes a huge difference. [5th grade]

> Right now, the requirement at our school is that we do writing in the reading/writing block, and the writing goes more with our reading curriculum. I can't say, 'well we are going to do science then'—they break students into smaller teams. If I had my whole class, I probably would spend more time on the science and do more of my writing then. [4th grade teacher]

The Lead Science Writing Teachers shared with us a number of observations about directions they would like to see the program go.

They feel it is very helpful, after the initial writing classes, for teachers to continue meeting to discuss their use of the writing strategies for the notebooks. Currently only the Lead Science Writing Teachers have formal opportunities of this kind. The LSWTs would like to facilitate such conversations among their colleagues:

> Beyond the writing classes, having ongoing grade-level meetings where I can go and talk with other 4th grade teachers is so valuable! I feel like, okay, everyone is having a hard time getting their kids to write this explanation, so it can't just be the writing, it must also be their understanding. That kind of helps, because it helps me revamp what I am doing, and we can work together to come up with some ways of helping both areas, the writing and the science. [4th grade teacher]

> For people who are working in one unit, it is helpful to sit down and talk to people about how it is going, how they are using the writing, the classroom management things they do, how do they get that concept across. We lead teachers get to do this, but there are five of us, so only five 1st grade teachers in the whole district are getting a lot out of the writing because they meet once a month. But if you had somebody at your school who could help with the conversations, and you were bringing in your science notebooks and saying, 'we are going to look for these things, and discuss with your teams how you can improve the writing,' or something like that. It seems to me one person from each school could sort of organize it once a month or six weeks during staff meetings–let's sit at these round tables and see what other people are doing, see other ideas, so then you know, wow, maybe my kids can do more, or maybe this idea worked well with that person. [1st grade teacher]

During the scoring sessions for the notebook study, the lead teachers told us that they had never before read through *entire* notebooks and discussed them with other teachers, nor had they had many occasions to read notebooks from classrooms other than their own. They found this opportunity tremendously helpful because they could gain different insights into both teaching and learning. Further, they said that the rubric used for this study has potential to support professional conversation beyond the initial writing classes:

> I think if we got together teachers who have been using the notebooks for a year or two, and we said, 'bring in the notebooks that you have' and we will read the whole notebooks like we are doing here, and then say, 'okay what do you think would help us to get kids from 2's to 4's next time?' And we just sit at a table like this and discuss it– what does this kid have in his notebook that is a 4? I think it would increase the rigor, what we expect from our kids. [1st grade teacher]

Some of the more experienced Lead Science Writing Teachers would also like to see ongoing development of the writing curriculum itself. They believe that the current version under-emphasizes inquiry in the form of student-generated questioning and student-designed investigation:

> Sometimes I give them the focus question for the day and sometimes I don't. I like the idea of a focus question, but then sometimes I am feeling like it really isn't inquiry-based when you give them the questions. I really had a switch in my thinking after I got back from the Exploratorium workshop on inquiry. Before the Exploratorium, I was using the kits just like they were, but now I am feeling the kits are saying to the kids, 'this is the experiment, do this experiment, what did you find out?' rather than being real true inquiry where the kids generate the questions. Even when I asked them to end an entry with a question, we didn't do anything with them. Now I want the kids to design their own investigations, tell what they predict will happen and why, and see what they find out. [5th grade teacher]

> We are trying to get the kids to develop their own investigations and shifting them to this more investigatorial way of thinking, and I don't see right now the writing going there. I don't know if the kids don't have the skills yet, or we don't have the time to develop the skills so that they can go there. [3rd grade teacher]

They realize, though, that making this change will be a challenge because of the nature of the science curriculum to which the writing curriculum is tied. To foster more student inquiry, the teachers would need even more time than they have now:

> Unfortunately, it is the units themselves. There is a lot of just guided inquiry. There are a few situations where the kits can open up to extensions if the teacher has the time to have students designing their own experiments and that kind of thing. But you have to really be dedicated and have a lot of time on your hands to pull that off. It's unfortunate, because we want kids to go there, but we simply don't have the time. [3rd grade teacher]

III. FINDINGS FROM THE STUDY OF STUDENT NOTEBOOKS

Since we began studying the Expository Writing and Science Program, we have been impressed by the kinds of messages summarized above, especially teachers' strong beliefs that the professional development gives teachers approaches and curriculum that are actually usable, and that it supports student learning in ways that they and the district value. When we directly observed workshops and examined the written materials, we observed that the approach is highly structured and consistent across units and grade levels. This gave us to believe that if we read a sizable sample of student notebooks, we could observe patterns of work that would reflect teachers' use of this structured approach. It thus seemed both desirable and feasible to carry out a study of the student notebooks. The notebooks, we surmised, could serve as a window onto the program, enabling us to verify the extent to which and the ways in which students' work reflects implementation of the program in classroom practice. A study of the notebooks would also help us assess the degree of progress students are making toward competence with respect to the program's goals.

Design of the Notebook Study

The sample

The notebook study involved 150 notebooks drawn from 15 classrooms in 13 elementary schools. Of the 15 classrooms, four were 1st grade, five were 3rd grade, and six were 5th grade.

From each of the 15 classrooms, we drew a sample of 10 notebook from the full class set; this sample was randomly drawn but was stratified to reflect the actual proportion of formally designated ESL and Special Education students in the class.[2]

The schools

In selecting the schools, we wanted a sample that reflected the full range of student demographics in the district, but we also purposefully over-sampled for schools with higher concentrations of poverty and ELL in order to explore benefits to traditionally low-scoring schools.

[2] There were often more students whose first language is not English but who were not identified as qualifying for special services; these other ELL students we counted as "regular" students.

The following table shows characteristics of the sample schools compared to the district average for K-5 schools on poverty and LEP status. Of the 13 schools, seven have higher than average proportions of both LEP students and students who qualify for free or reduced lunch; these seven are shown in bold italics on the chart. Two additional schools have higher than average numbers of LEP students.

Table 1.
Poverty and LEP Status in Sample Schools for the Notebook Study

School	Number Enrolled	Free or Reduced Lunch	Limited English Proficiency
District Average	**335**	**44.5%**	**17.5%**
School "A"	***386***	***59.1%***	***33.4%***
B	***404***	***78.7%***	***34.4%***
C	***413***	***86.4%***	***30.3%***
D	***353***	***56.1%***	***22.9%***
E	295	22.4%	3.4%
F	***442***	***64.9%***	***28.5%***
G	372	29.0%	22.0%
H	***430***	***60.9%***	***37.0%***
I	267	8.6%	0.0%
J	***312***	***75.0%***	***24.4%***
K	298	19.1%	3.4%
L	307	34.2%	18.9%
M	368	12.8%	2.4%

From Seattle district, October 2001.

The teachers

In selecting the teachers, we selected those who had participated in at least three of the four available workshops for their grade level (the introductory one and at least two focusing on specific units), and whom the program leaders considered to be active participants in and implementers of the program. That is, we did not select participants randomly to ascertain whether they were implementing the program; rather, we selected implementers to examine student work related to the program. For practical reasons, two other factors were involved in selection of teachers: they were willing to offer their notebooks for this study, and they had full class sets of notebooks available (science notebooks are very popular take-home products).

We asked these teachers to fill out a background survey to get information about their experience and their self-identified comfort level with and use of the program

materials.[3] Of the 15 teachers, seven have been teaching five years or less. Eight have been teaching hands-on science using district adopted kits for three years or less. All 15 are responsible for teaching all four core subjects of language arts, mathematics, social studies, and science, and some also teach other subjects. All 15 of them rated the value of the writing program's workshops highly; 13 said they were more valuable than any other professional development in writing they have received, and 11 said they were more valuable than any other professional development in science. For 10 of them, the writing approaches they learned in the program were completely new, and for the other five they were partly new. Fourteen of the teachers believe they are consistently and fully putting into practice the approaches they learned in the program. They believe that the strongest support for their implementation of the program is their belief in its value, and the greatest barrier is time allotted to the teaching of science.

The notebooks

It is important to note that this was a completely natural sample and naturalistic study. That is, neither the students nor the teachers knew ahead of time that the notebooks would be used for a purpose other than ordinary classroom instruction. They were not written for an external assessment nor were they prepared in any standardized way; rather, they reflect the types of use, and variation in use, natural to their school and classroom contexts.

The raters and scoring sessions

We held two separate reading/scoring sessions with two groups of raters.

The first session, held in Seattle in May 2002, involved Lead Science Writing Teachers in the program, whom we refer to below as the "Seattle raters." These are teachers who have helped develop and field-test program materials within their grade levels. They are immersed in the program goals and the standards, and their role was to bring those program standards and their expert-practitioner perspective to assessment of and reflection on the notebooks. They played this role by verifying that the scoring criteria reflected the program standards, and also by verifying that particular characteristics of student work matched the different levels of development on the 4-point rubric. When these raters assigned scores, they were thus assessing the extent to which the type and quality of work in the notebook matched program standards.

The second session, held in Inverness in July 2002, involved a panel of six experts who are independent of the Seattle program. The panel included classroom practitioners with expertise in writing and science, professional development leaders in science reform and national standards, and researchers of professional development and

[3] The full results of the survey are included in Appendix II.

curriculum reform programs in elementary science education and writing. This panel—whom we refer to as the "independent raters"—brought to the study accumulated background knowledge and experience that reflect the broader standards and perspectives of the field. Their role in examining the student work was to bring this broader perspective to it, that is, to assess the extent to which the notebooks contained evidence that students are being introduced to, and are developing, skills and knowledge that are regarded by the national reform community as being valuable. In this way, the panel used the notebooks to make judgments about the goals and values of the program.

For the Seattle session, all 150 papers were read and scored. For the Inverness session, a sub-sample of 67 (45%) were read and scored. All readers were blind with respect to school and student identification, as well as ESL or Special Education designation. The independent panel members were blind to the Seattle raters' scores.

The ratings criteria

Our purpose in developing a scoring guide for the science notebooks was to form criteria that reflected the aims and standards of the program, and with these as a foundation, to develop a scoring scale (rubric) that could be normed against the sample of notebooks within each of the three grade levels we sampled. We first reviewed documents provided by the program that specified the goals and standards to which they anchored the supplementary writing curriculum,[4] the professional development classes, and their internal guides for formative assessment of student work.

These materials generated the three major criteria against which the student work would be assessed at each grade level:

I. Conceptual understanding, i.e., evidence of student understanding of the core science concepts to which they were introduced.

II. Scientific thinking, i.e., evidence of students' use of skills, processes, thinking, and discipline perspectives that are fundamental to scientific inquiry.

III. Expository writing, i.e., evidence that students can present ideas and science content in a well-developed and organized way with accurate use of vocabulary.[5]

[4] According to program leaders, the program goals and criteria reflect science standards important in Washington, as well as standards for writing that appear in the WASL and Seattle's Direct Writing Assessment. Program documents reflect these standards.

[5] The fully defined criteria and scaled scoring rubric are included in Appendix III.

We then read in an exploratory manner a pilot sample of science notebooks that reflected the range of student work existing naturally in the schools within each grade level range. With these as a reference, we drafted a scoring rubric that spread each of the three criteria across a 4-point scale—from *limited*, to *developing*, to *adequate*, to *full*.

Betsy Fulwiler, coordinator of the program, then reviewed the full definitions of the three criteria and the 4-point scales to verify that they reflected the program's goals and standards.[6] During the three-day reading/scoring session held in Seattle, we further validated the rubric as being consistent with the program by test-scoring sub-samples papers at each grade level with the Lead Science Writing Teachers. They verified that the definitions and rubric accurately reflected the program's goals and standards.[7]

During the reading sessions, each rater read the entire body of work in a notebook, and then assigned each notebook three ratings, one for Conceptual Understanding, one for Scientific Thinking, and one for Expository Writing. Each of the three ratings ranged from 1 to 4, as shown in the rubric. We then assigned each notebook a total score, simply the sum of the three individual scores. The range of total scores runs from a minimum of 3 to 12.[8]

In this section of the report we present results both for individual criteria (range of 1-4 for each) and for total scores (range of 3-12). The following table reviews the levels of development and competence within these ranges:

Table 2.
Ranges of competence and skill reflected in the scoring guide[9]

	Limited	Developing	Adequate	Full
Scale for each criterion	1	2	3	4
Range for total scores	3	6	9	12

[6] Betsy Fulwiler, along with Elaine Woo, offered a tremendous amount of assistance with the many time-consuming steps involved in setting up the study and shipping quantities of notebooks to various places they needed to go. They did this with scrupulous integrity and in a true spirit of inquiry.
[7] In fact, the teachers believe the scoring guide is so comprehensive and useful that they wish to incorporate it into their professional development activities.
[8] We reserved the score of "0" for a notebook that contained no evidence of the criteria. Of the 150 in the sample, only one received a 0. Thus we report results for n=149 notebooks.
[9] In Appendix IV we include a number of annotated excerpts from student notebooks to illustrate what student work looks like at different levels of these criteria.

Results of Notebook Ratings

Below we discuss findings related to these three areas:

A. Knowledge and skill development evident in the student notebooks, from the perspective of the program

B. Correlations between notebook ratings, school demographics, and WASL scores

C. Knowledge and skill development evident in the student notebooks, from the independent perspective

D. Reflections on the program's contributions to teaching and learning, from the independent perspective

A. Ratings of Notebooks by Program Leaders

These results represent program's perspective on the extent to which the program's goals and strategies are reflected in students' work, and evidence of students' competence in the knowledge and skills related to their program's standards.

Overall finding: The notebooks are consistently used in ways that reflect attention to program goals. The quality of student work overall shows notable progress toward program standards on all three criteria, with notebooks of 3rd and 5th graders slightly closer to the *adequate* level than 1st grade notebooks.

Ratings on notebooks of students with identified special needs are roughly half a level lower on average than those of other students, with differences between student groups smallest in 1st grade and largest in 5th grade. However, for these students there is also considerable evidence of student progress toward program goals.

Overall total notebook rating

For all notebooks across all three grade levels, the average total score assigned by the Seattle group was **8.3**. This total is 2.3 points above the *developing* level of 6, and .7 point below the *adequate* level of 9.

This result suggests, at the most fundamental level, that the notebooks in the participating teachers' classrooms are, in fact, consistently used in ways that clearly

reflect program purposes: they show clear and consistent evidence of work toward the standards. The results also suggest that the student notebooks overall reflect a notable degree of progress toward competence with respect to program standards and criteria, with room for continued improvement.

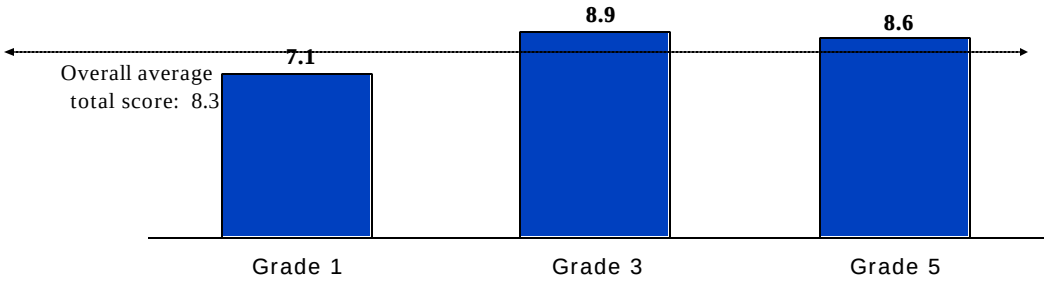
When we asked the Seattle raters to reflect on the quality of work in notebooks at the end of the scoring sessions, a number of reflected on how far the program had come in a short time and also on its potential for supporting even further development. One teacher put it this way:

> Even though some of the notebooks that we saw today were just, wow, phenomenal, I still can see the room for improvement. I know for a fact that my ability to get something good going in my students' notebooks in the past year has increased dramatically. I am really encouraged about what I am going to be able to do in the next year.

Total notebook ratings compared by grade level

On average, 1st grade students were rated at 7.1, which is 1.2 points lower than the overall average of 8.3, and somewhat closer to the *developing* level of 6 than to the *adequate* level of 9.[10] Notebooks of 3rd graders received an average rating of 8.9, and 5th grade notebooks an average rating of 8.6, somewhat higher than the 1st grade set and close to the *adequate* level of 9.

Figure 1.
Seattle Raters
Comparison of total scores by grade



A possible contributor to the variation between the 1st grade sample and the 3rd and 5th grades may be that the 1st grade sample included fully 50% identified special needs students (Special Education and LEP), whereas the other two grades' samples included

[10] Notebooks of different grades were not compared to one another, but only compared to (normed against) notebooks within their own grade level. Thus, more 1st grade notebooks were rated at the *developing* level than at the *adequate* level within the range existing in the 1st grade sample.

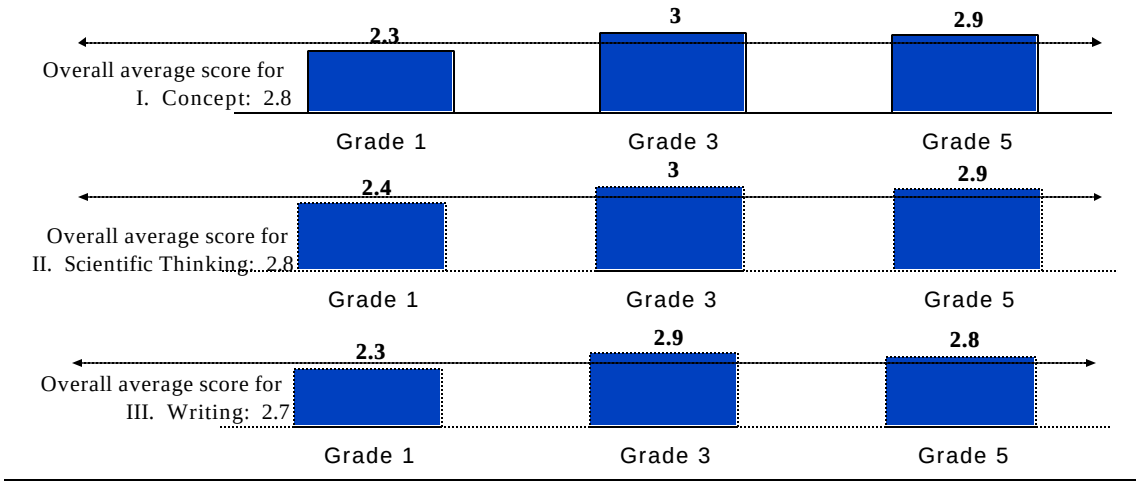
20%. Additionally, all four of the schools where 1st grade samples came from have higher levels of poverty than average in the district. In contrast, in the schools where the 3rd grade samples were drawn, only two of the five schools have poverty rates above the district average, and in the schools where the 5th grade samples were drawn, only two of the five schools have poverty rates above the district average. Below, we display correlations between these demographic characteristics and notebook ratings.

Ratings of each criterion by grade level

Across all grades, the average scores for the three criteria are 2.8, 2.8, and 2.7 respectively on the 4-point scale. Within each grade, scores for each criterion are either identical or .1 point apart.

There are no statistically significant differences between scores for the different criteria. These results suggest that these three knowledge and skill areas go hand-in-hand with respect to student development.

Figure 2.
Seattle raters
Comparisons of scores for each criteria by grade



A closer look at results for special needs students in the sample

Of the 149 students whose notebooks were rated, 42 (28%) are formally identified for either Special Education or English Language Learner services.[11] We compared ratings on these 42 notebooks to ratings on all others. We also made comparisons within each grade level. In the sample of 1st grade notebooks, fully half of the students (20 of 40)

[11] Those who rated the notebooks were blind to these designations.

were identified as having special needs.[12] In 3rd grade, ELL or Special Education students comprised 20% of the sample, and in 5th grade they comprised 24%.

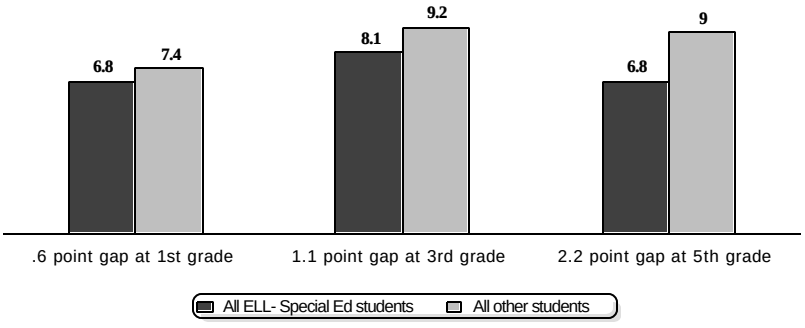
The result was that, overall, ratings for special needs students averaged **1.7** points lower than those of other students. The gap between groups is narrowest in 1st grade and widest in 5th grade.

The notebooks of the 42 special needs students averaged 7.1 on the 3-12 point range of total notebook scores, placing their average level of competence against program standards at 1.1 points above developing (6). Scores of other notebooks averaged 8.8, which is just .2 below the adequate level (9).

When we looked at differences at each grade level, we observed a trend. Although the proportion of special needs students is greatest in the 1st grade (half the sample), the gap of .6 point between the two groups is smallest at that grade. The gap between the two groups widens until it is 2.2 points at the 5th grade level.

The figure below shows the comparisons for all grades.

Figure 3.
Comparison of total notebook scores of special needs students and other students for each grade



We do not have a statistical explanation for this trend. The group of independent raters, however, developed a general impression during their reading of notebooks that there seemed to be a wider gap between low and high scores in the 5th grade sample than in the other two grades. They speculated at the time that some of the LEP or Special Education students may be left "farther behind" by the time they reach 5th

[12] The one non-scored (blank) notebook was not from a special needs student.

grade—perhaps offered less in the way of individualized support in class, or perhaps pulled out of class more frequently. We explore this more fully later in the report.

B. Relationship of Notebook Ratings to School-level Characteristics

For purposes of exploration, we conducted two other comparisons. First, we compared total notebook scores with the student poverty levels and ELL concentrations in the 13 schools. Next, we did a three-way comparison, looking at WASL 4th grade writing scores and also notebook writing scores, compared to school poverty and ELL rates.

We want to caution that these results are only suggestive, given the very small number of notebooks read in each school. We carried out these analyses more for purposes of provoking questions than for drawing conclusions. Taken alongside the reflections of those who rated the notebooks, they may be suggestive of the program's potential to make a contribution to schools' efforts to narrow the gap in outcomes for lower SES students.

Overall finding: **For both notebook scores and WASL scores there is a general trend toward lower scores in schools with higher populations of at-risk students. However, there are several schools in the notebook study where results counter this trend, suggesting there is potential for the program to make a contribution to more equitable achievement.**

Ratings by school by student poverty rate and LEP concentration

As we noted earlier, our sample of schools reflects a wide range of student demographics, but we purposefully over-sampled schools with higher than average poverty and ELL rates. Seven of the 13 schools in the sample have poverty rates higher than the average of 44% in the Seattle schools. In two of those seven schools, two class sets of notebooks were pulled for the study (in one school, two 5th grade sets, in the other school one 3rd and one 5th grade set.); thus, nine of the 15 class sets of notebooks came from schools that have a higher than average poverty rate. Of the 13 schools in the sample, nine have concentrations of ELL students higher than the Seattle average of 17.5%. In two of these nine schools, two classrooms were sampled, so 11 of the 15 classrooms were in schools with higher than average ELL populations.

We found that there is a general trend toward higher scores for schools with lower rates of poverty and ELL populations and lower scores for schools with higher rates of these demographics. However, there is a cluster of five schools that have higher-than-

average rates of both populations but whose scores range between 7.4 and 9.0, less than one point different from the overall average of 8.3.

We display these results in the two graphs below. We have circled the points for the five schools.

Figure 4.
Total notebook scores compared to % of students in school qualifying for free or reduced lunch

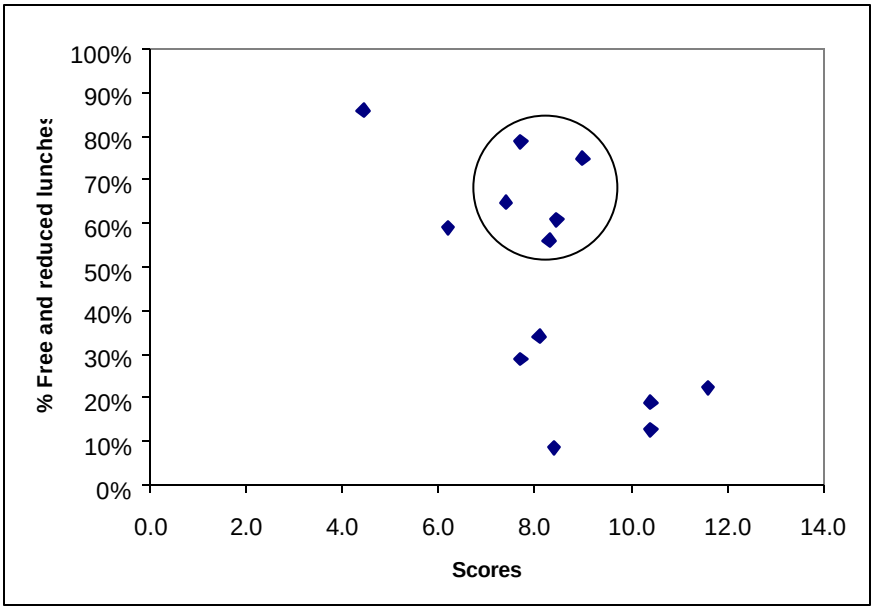
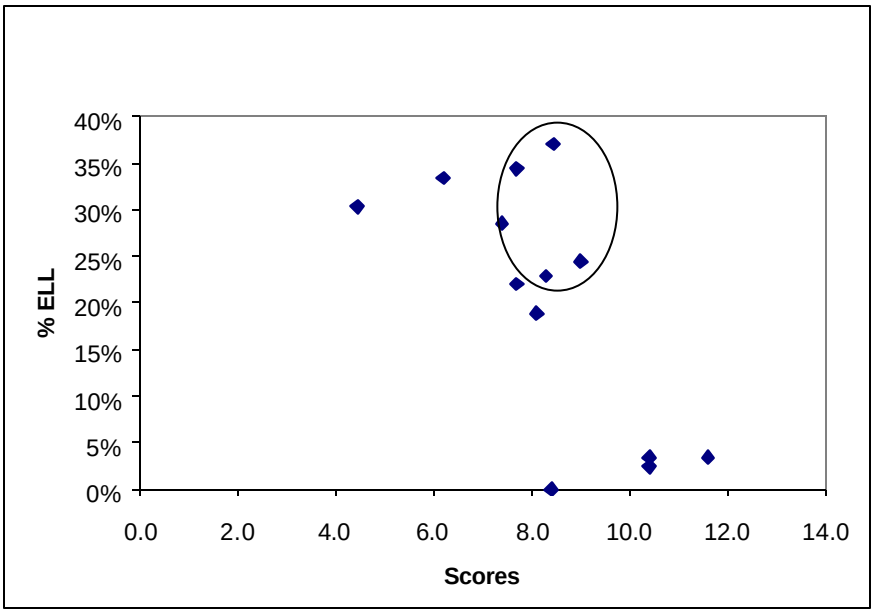


Figure 5.
Total notebook scores compared to
% of LEP students in school



This pattern is provocative because it suggests that there may be potential for a "leveling of the playing field" in schools with at-risk student populations.

A total of seven of the 15 teachers whose notebooks were sampled teach in these five schools. To discern whether there were any differences between these teachers' implementation of the program and that of the other eight teachers, we compared their responses on key items on our background survey. The group of seven teachers in these five schools rated their implementation of the program .6 higher on a 5-point scale; and they rated the degree to which they have increased the amount of writing students do at .5 higher on the 5-point scale. Because of the small *n* in this group and the fact that nearly all the responses were at the upper end of the scale, though, these differences are not statistically significant.

Comparison of correlations between school characteristics, 4th grade writing WASL scores, and notebook ratings for writing

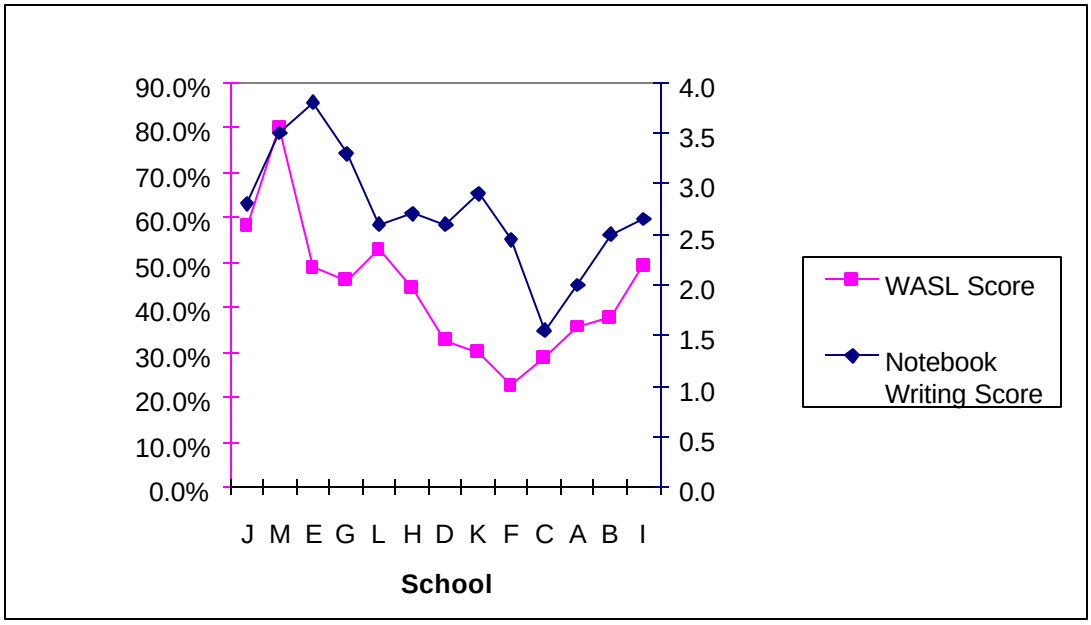
We then compared notebook ratings on the 4-point-scaled writing criterion for the notebook study against 4th grade WASL writing scores for the 15 sample schools, again by the two demographic characteristics of ELL concentration and poverty rates. We were interested generally in the extent to which WASL writing scores correlate with demographic characteristics. More specifically, we wanted to see what relationships

there were between WASL scores and notebook writing scores for the five trend-breaking schools we refer to above. On the graphs below, these five schools are identified as B, D, F, I, and K.

We observed that both WASL scores and notebook ratings show a rough trend toward lower scores in schools with higher LEP concentration and poverty levels, though there are individual schools in this sample that appear to counter that trend. In several schools, notebook scores show a greater tendency than WASL scores to counter this trend; that is, there is less of a tendency for notebook scores to be correlated with school ELL and poverty. Again, this suggests that the notebook program may have potential to contribute to a narrowing of the gap between at-risk and more advantaged groups.

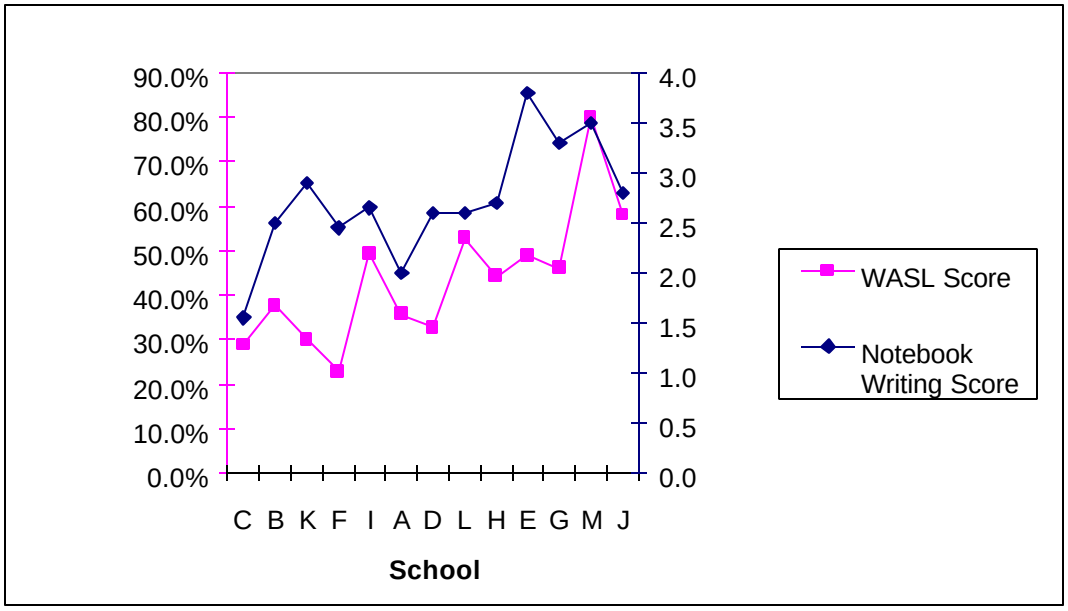
The graph below shows WASL and notebook scores arrayed by ELL concentration. There is a sizable gap between the two curves at schools D, K, and F; this gap suggests a weaker relationship between notebook ratings and LEP concentration than that of WASL scores. For schools B and I, which have the heaviest concentrations of ELL students, there are smaller gaps between the two lines, with both WASL scores and notebook ratings countering the general trend toward lower scores for higher LEP.

Figure 6.
4th Grade Writing WASL Scores 2000-01
% Who Met Standards by LEP Status
Lowest to Highest LEP Schools



The graph below shows WASL and notebook scores arrayed by concentration of students who qualify for free or reduced lunch. Both curves show a similar rough trend toward higher scores in schools with higher wealth. However, there are sizable gaps between notebook ratings and WASL scores for four of the five high-poverty schools of interest: B, K, F, and D. For these four schools there is a weaker correlation between wealth and scores for the notebook ratings than for the WASL scores.

Figure 7.
4th Grade Writing WASL Scores 2000-2001
% Who Met Standard by Poverty Rate
Poorest to Richest Schools



C. Results of the Scoring of Notebooks by Independent Raters

The total number of notebooks scored by both the Seattle raters and the independent raters was 67, or 45% of the total sample of 149 scorable notebooks. We made comparisons of scores given by the two groups, using the 67 notebooks that had been read by both Seattle and independent raters.

Overall finding:

The independent raters found consistent evidence in the sample of notebooks that students are doing work related to the program's goals, and that students are developing knowledge and skills that are valued by the field. Their average total rating was 7.9, indicating their judgment that students' overall level of competence is moving toward the *adequate* level.

Where there are differences between the ratings given by the Seattle raters and the independent raters, the trend is for the scores of independent raters to be lower than those of the Seattle raters. The magnitude of the difference averages less than one point on the scale of total scores, and less than a half point for each criterion.

The difference for the Scientific Thinking criterion is slightly larger and is statistically significant. The independent raters thus have a somewhat different standard than the program with respect to high-level competence in scientific thinking.

Overall comparison of independent raters' scores with Seattle raters' scores

We began by comparing the total notebook scores given by both groups across all grade levels.

Overall, the independent raters' average score was **.8** point lower than the Seattle raters' scores on the 3-12 point scale for total scores. Overall, the independent raters' scores are closer to the *adequate* level than they are to *developing* level (two on each criterion, six on total score), even though they are lower than the average scores of the Seattle raters. The magnitude of difference in overall scores is thus quite small.

The independent raters' scores were .2 to .4 point lower on each of the three criteria. One difference—that for the Scientific Thinking criterion—was statistically significant, meaning it is not due to chance.

The following table displays the comparisons of all scores:

Table 3.
Comparison of scores of Seattle raters and Independent raters

	Average Total score	Average for I. Conceptual Understanding	Average for II. Scientific Thinking	Average for III. Expository Writing
Seattle raters	8.7	2.9	3.0	2.9
Independent raters	7.9	2.6	2.6*	2.7

** significant at p=.01*

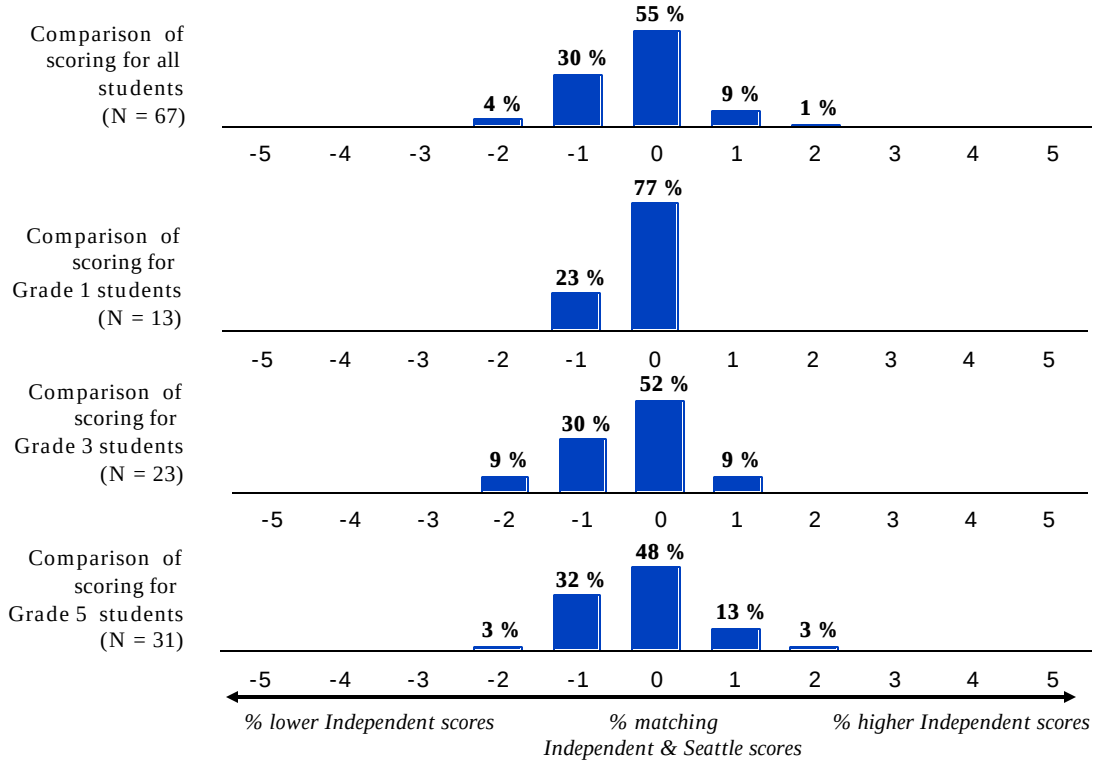
Comparison of each matched score on each notebook

We also analyzed how often the Seattle and independent raters matched scores exactly on each of the three criteria for each notebook, and measured the size of the difference when scores did not match. These breakdowns help reveal more detailed patterns in the variations of ratings.

We found that there were more exact matches than differences, and most differences were one point in magnitude. The differences between the two groups are somewhat more pronounced in 3rd grade, especially for the criterion Scientific Thinking.

The first graph displays the comparison for the Conceptual Understanding criterion. The spread of differences between the two groups' scores grows slightly wider from 1st grade to 5th grade.

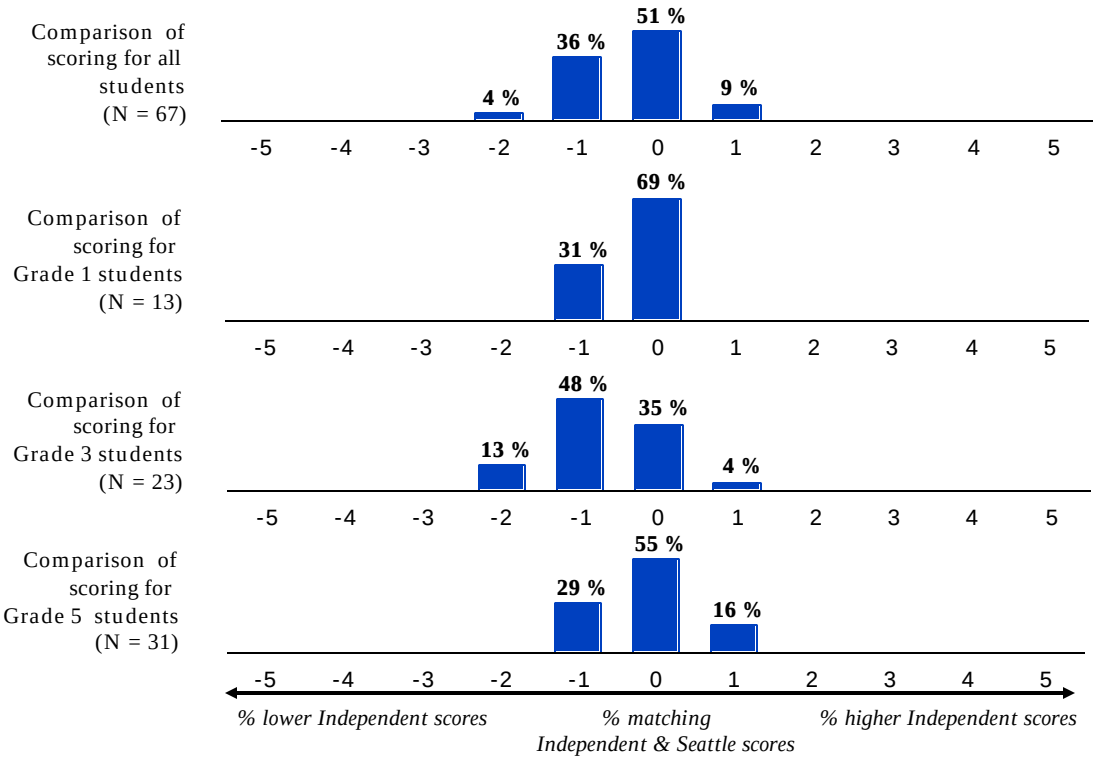
Figure 8.
I. Conceptual Understanding:
Difference between scores by Independent and Seattle raters



This graph represents the percentage of notebooks to which the Independent rater either assigned a lower score for Conceptual Understanding (i.e., gave a rating 1 to 5 points lower than the Seattle reviewer), gave the same score for Conceptual Understanding as the Seattle rater (i.e., "0" difference in points), or gave a higher score than the Seattle reviewer (i.e., 1 to 5 points higher).

In the next graph, for Scientific Thinking, there is a pattern of greater spread overall. In 3rd grade, there are fewer exact matches than there are 1-point differences.

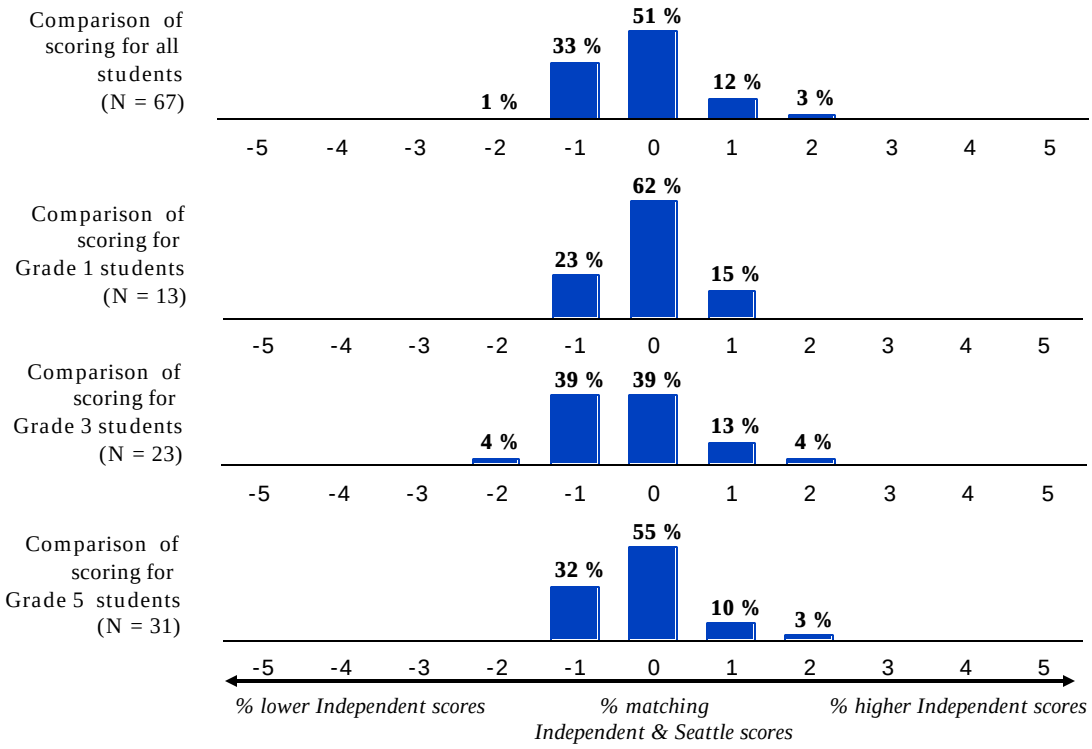
Figure 9.
II. Scientific Thinking:
Difference between scores by Independent and Seattle raters



This graph represents the percentage of notebooks to which the Independent rater either assigned a lower score for Scientific Thinking (i.e., gave a rating 1 to 5 points lower than the Seattle reviewer), gave the same score for Scientific Thinking as the Seattle rater (i.e., "0" difference in points), or gave a higher score than the Seattle reviewer (i.e., 1 to 5 points higher).

The following graph shows the breakdown of score differences for the Expository Writing criterion. For this criterion, as with Scientific Thinking, we observe that the spread is widest in the 3rd grade.

Figure 10.
III. Expository Writing:
Difference between scores by Independent and Seattle raters



This graph represents the percentage of notebooks to which the Independent rater either assigned a lower score for Expository Writing (i.e., gave a rating 1 to 5 points lower than the Seattle reviewer), gave the same score for Expository Writing as the Seattle rater (i.e., "0" difference in points), or gave a higher score than the Seattle reviewer (i.e., 1 to 5 points higher).

Independent raters' perspectives on the criteria

During the scoring session with the independent raters, we asked them to offer their own perspectives to the definitions of the criteria and to student work that reflected various levels on the rubric. Their comments help explain the difference that appears in the two groups' ratings for scientific thinking.

In these comments, independent raters offer their vision of "real" inquiry as scientific thinking that evolves in the context of student-generated questions and investigations:

> There may be few 4's in scientific thinking. For example, as I look at this notebook, are there 'accurate and full observations with complete records'? Yes, that would be a 4. Does the student 'use evidence to support explanations'? Yes. But is there 'a questioning stance' related to phenomena? to evidence? No. I don't see that. I saw a few little questions in the notebook—"I wonder what is going to happen to my plant"—but

> that is not what I mean by real questioning. There is no evidence of designing investigations to test questions. There was no evidence of the student moving on toward real inquiry, real scientific thinking.
>
> Self reflectivity, to me, would be the highest level–where someone is thinking out loud about their learning and they are learning through their own experimentation. A notebook like that would say, 'I thought this, but now I think this, because of this.'

In the following comment, an independent rater suggests that too little attention to student-generated inquiry and reflective scientific thinking may put a "ceiling" on what is possible for students, and that a sparing use of the "4" score would be a way to indicate that the outside group is applying a different standard to the notebooks:

> Our way of indicating the presence of a "ceiling" on student inquiry through the use of this rubric is to make some decision that if we see evidence in a notebook of only one or two of these aspects of what the rubric calls 'skillful and purposeful scientific thinking,' then it doesn't merit a 4.

The independent raters also speculated that science units themselves varied in the extent to which they invited real scientific thinking, and that this obviously affected what students wrote in their notebooks. One person put it this way:

> It was difficult to rate scientific thinking sometimes because the nature of the units is that they sometimes don't ask the students to do anything beyond observe and report. You can't ding the kids because they didn't make inferences if the opportunity to create an inference wasn't provided.

In the section below, we summarize independent group's further reflections on the notebooks, their benefits to students and teachers, and their thoughts about the program.

D. Independent Raters' Reflections on the Student Work, Its Benefits, and Implications for the Program

We interviewed the independent raters during and after their scoring of the notebooks. We asked them to make observations about the student work they had examined, and we also asked them to posit more general assessments and inferences about the program: its benefits to teachers and students, its potential to make an important contribution in Seattle, and the challenges they believe the program faces.

Overall finding:

The independent panel views the Expository Writing in Science program as making an important contribution to student learning in both writing and science. They believe approach is especially effective in supporting the language and learning development of ESL students and others struggling with literacy in English.

When considered in relation to other elementary science programs with which the independent panel is familiar, the Expository Writing and Science Notebooks Program stands for its deliberate approach to both curriculum and professional development, and for its outcomes in terms of the student work produced.

In its next phase, the independent raters believe the program should continue providing professional development to large numbers of teachers, as well as working to increase its emphasis on supporting student-generated inquiry.

The quality of work in notebooks and benefits to students

The independent raters' views about the benefits of the program are quite consistent with those of the participating teachers and lead teachers. Most fundamentally, the independent raters saw considerable value in the scientific skills and habits of the discipline that the students evidenced in their notebooks.

> This structure of writing seems to help them in their scientific thinking. They are internalizing the idea that you don't just come up with an idea out of the air, but you think scientifically–I noticed this, therefore I learned.

The raters saw considerable evidence that the approach gave ESL students and others who are struggling with language development—especially younger students—a valuable springboard into written English and an important scaffold of language and logic on which to build their learning of science.

> A number of notebooks were clearly from kids that weren't speaking English well at all. They might start out their notebook barely able to put a sentence together because of the language issue, and by the end, they were much better able to do writing. They were learning to write, learning to communicate in the context of doing science.
>
> They are being given a structure that allows them to learn a language that is not their first language, and to demonstrate what it is that they know and learn in science.
>
> Kids don't walk in being able to think and record and write in logical step-by-step-by-step ways. They don't come knowing that there are signal words, or knowing there are

> transition words, or knowing that you are necessarily supposed to do your evidence and then draw your conclusions from your evidence. That needs to be taught.

The independent raters saw that because these students were given access to language and logical structures, they were able to provide teachers with evidence of their learning in science that they could not have done otherwise. This in turn seemed to afford better means of assessing student understanding:

> The child is not ready to generate the sentences in English but they can follow the model. And since they have done the science experiment and they are filling in the data, it allows the teacher to see what the child knows from having done the experiment, rather than for the teacher to see what the child can't do. We know this kid can't do this independently, so why put a writing task before them that is going to frustrate them and probably make them feel badly and perhaps not want to do science. At the same time, it is teaching the structure of the language and the protocol for this kind of thinking and writing.

The independent raters saw, further, that some students could use the writing and thinking structures as tools to support "their own thinking" rather than being constrained by them. They saw more instances of this in the higher grades but saw some at every grade level. For these students, the program's approach helped develop valuable skills while still inviting ownership of the scientific work.

About a 3rd grader:

> Once it is taught and it becomes a tool that they can pull out of their pocket and use when they need to do that kind of writing, then I think that is where they move toward internalizing it. For some 3rd graders it seemed very lock-step, but we saw some who, when they were more engaged, actually were able to call upon those frames and use them in more authentic ways, where they seemed to own them and were comfortable with them, and there was more of their own thinking.

About a 5th grader:

> He was using the scientific vocabulary in a very natural way and very confidently. He started a lot of his sentences with I-'I noticed that this happened, 'I think this because'– he has the personal connection to what he is doing and observing, and he is having fun. It is though he has internalized it at this point. To me that would be the purpose of giving that kind of a pattern, you want them to be able to do this kind of writing.

On the following page we offer an excerpt from a 1st grade notebook (link coming soon to sample of student notebook) that the independent raters identified as a good example of how the approach can support the development of both writing and science learning for a student in the early stages of literacy development. There are three written entries on the page, all several weeks apart. Together, they show the child moving from heavily scaffolded to more independent work. The entry on the left, done in October on

the Balls and Ramps unit, shows that the student was copying—by tracing over each letter—a series of sentences setting up a comparison between two balls, and then writing in on a blank space provided the one or two-word result of the trial. In the middle entry, done in early February on the Weather unit, the student has transcribed, in its entirety, a paragraph the class composed together.[13] In the right-hand entry, done in mid-March, the student enters her own independent composition.[14]

The notebook received a total score of 6, i.e., a score of 2 on each of the three criteria, indicating observable progress toward program standards. The student records observations from guided investigation in all three entries. In the last one, where the student records the air temperature, the student also formulates and communicates a reasonable conclusion from the evidence.

[13] This same paragraph appears in each notebook from that class on that date.
[14] Students in the same class wrote different reports, though all observed that the temperature outside was 40°.

The raters voiced a concern about a pattern they observed in the 5th grade sample—a pattern that is borne out in the range of scores for that grade. There were a number of notebooks where the students' work began at a limited level and did not show development over time. They wondered whether the writing curriculum designed for the 5th grade units may leave struggling students behind:

> If you have a 5th grader who has language or learning issues, it is not necessarily that this systematic approach to writing wouldn't work for them. However, if it is taught in the same way to all 30-odd kids in the class, if some students have problems with learning, the writing and the science is too many things to hold onto at one time. It is just too much without one-on-one support or other scaffolding. So those kids will just sit there and draw a cat down their margins.

At the same time, they were concerned that the same writing curriculum was overly constraining to those students who began at a high level of competence. This led the group to observe that some differentiation of approach may be more optimal for older students:

> I saw less growth for the children that are proficient writers. I wonder if they could sometimes offer different options, you know–'in your notebook you can do option A, option B, option C'–to go beyond and do something differently, or not have to do the same thing lesson after lesson after lesson.

The independent raters believe the writing curriculum could be designed with a more developmental purpose in mind, so that it leaves no students behind but also leads students to increasingly more difficult science ideas and independent thinking and writing.

Comparing Seattle notebooks to "typical" elementary science notebooks

The panel of independent raters brought to this study extensive experience both in leading and in evaluating elementary science programs around the nation, and in teaching writing in the context of science at the elementary level. We asked them to make implicit comparisons between the Seattle notebooks and "typical" elementary science notebooks.

The raters said that the Seattle notebooks stand out from notebooks they have seen in other programs because they contain a greater amount of student writing overall, they reflect a much more deliberate and systematic approach to developing students' writing and science skills, and they offer students greater opportunities than typical science notebooks to formulate and express their ideas in science. Below, we highlight observations specific to each grade level:

First grade:

> There is a lot of writing here for a 1st grader, absolutely. A typical 1st grade written form is draw a picture, write a sentence. First grade paper usually has a big space that is blank and then a few little lines for writing, and of course that sets up an expectation that this is all you are capable of doing in 1st grade. But there is a lot of text in these notebooks. It is sequenced, they are using transition words that I am presuming that they have been taught, and they make sense. They are using lots of scientific vocabulary, collecting data and then making a statement based on it–'My data shows me, my graph shows me.' It is extremely structured, but they are able to use it to get a lot of language out and onto the paper.

Third grade:

> You are not going to see this kind of writing in very many 3rd grade classrooms. We are getting something to really look at to see what the kids know. I think it is great.

Fifth grade:

> I think the Land and Water unit affords students the chance to really think about erosion and deposition and deltas and canyons in ways they never thought before. But countless times in classroom observations I haven't seen a journal where kids could actually write that. I could hear it in table talk, but when they are then told to write in their journal I haven't seen their thinking come through. So it strikes me that this student can write quite clearly and fully about it. It is not just table talk anymore–he can actually translate his thinking into writing on a page, and that is what really struck me about reading his journal.

Questions about the program's approach and suggestions for its next steps

As the raters studied the student work, they naturally made inferences about the program's aims and its values. They also raised questions and possibilities for ways the program could continue on a path of development that would benefit students, teachers, and the district more broadly.

The raters were impressed that the sample of notebooks showed such obvious evidence of instruction that related to the program's goals and aims. They inferred that the program has strong potential for wide-scale implementation through ongoing professional development and support:

> It seems like a natural for a large district because it is so step-by-step, with the writing curriculum laid out there with the science units. It is helping them teach the science, there's no doubt, I mean they are using it. And I think it's helping them give the kids some important learnings in writing. In large districts where you have so many teachers who aren't comfortable with teaching science, they're not prepared, I can see that this would help. It would definitely "raise the floor" of hands-on science teaching by giving

> teachers real strategies they can use. And it can also raise the floor for students because it gives kids a way to get started.

These raters also inferred that, thus far, the program was "making a choice" to prefer developing children's conceptual knowledge through guided inquiry over developing children's natural curiosity and intellectual power as inquirers. The raters also observed that because the notebook program is embedded in, and serving, the curriculum of the science program, there are reasons for the under-emphasis on student-generated inquiry.

> What I see here is science following someone else's pattern, someone else's thinking. A lot of it is what I would call guided inquiry. It is a guided progression through experiences that kids have that are going to lead them to some 'big ideas'–but there actually isn't very much inquiry in the kits. On top of that, the notebooks are structured in such a way that the cognitive, intellectual inquiry stuff that may be happening to kids as a secondary effect of the curriculum also isn't given an avenue or structure for exhibiting itself. There is an aspect of doing science that has to do with promoting children's natural curiosity and giving them opportunities to look at things in novel and exciting ways, so that what begins to emerge is an attitude about the world, and the child is curious and excited and filled with power about their relationship to all of that. To me as a science teacher, that is one of my primary goals in teaching kids science. I don't see that here. There seems to be a ceiling on what is possible here.

It is important to note, here, that the raters understood this concern to be about the science program as a whole—both the kit-based curriculum and the approach to writing that serves it. They found, for example, that the degree of student engagement and questioning found in the notebooks varied with the kits. This may have contributed, in part, to the lower ratings in 3rd grade:

> I found the 3rd grade more problematic, but again it could have been the science units they were doing. The 5th graders got to do a more engaging unit on Land and Water, and the 1st graders—who wouldn't have fun playing with Balls and Ramps? But the 3rd graders, many of the notebooks were about the Sound unit, and it is dry! They seemed rote in a way–"if it is bigger, it is lower." Once they got that rule down, the learning sort of stopped. Again, is it the writing model? Or is it the science kits?

Even as the group voiced this concern, they also suggested that it may be the notebooks program that can help lead the way in fostering more student inquiry within the elementary science program. One way is to invite Lead Science Writing Teachers to experiment with a wider range of prompts/focus questions:

> It might be interesting for some small group of those folks to think about scientific thinking in different ways and to play around with different kinds of prompts and see what it brings. If they could look at ten kid responses to prompt A in contrast to ten kid responses to prompt B, and prompt C, I think that would be very instructive for them. It might actually lead them to the bigger idea if it immediately kicks them back into 'what am I really trying to do here, what is really important?' It might allow them to let in a little

> bit more intellectual space. It might allow them to think, this is one way of achieving particular results, but there could be other ways of achieving other kinds of results.

Though the independent panel has a stronger concern than most participating Seattle teachers that the program may be limiting students by under-emphasizing student-centered inquiry, we do note that some of the Lead Science Writing Teachers hope the program can make greater strides in this direction.

In sum, the independent raters see the notebooks program as making an important contribution to student learning in writing and science. The program stands out from other elementary science programs in its deliberate approach and in its outcomes in terms of the student work produced. In its next phase, the raters hope program leaders can explore ways to increase the emphasis on supporting student-generated inquiry, while still providing professional development that supports implemented by a wide spectrum of teachers. One person summed it up this way:

> They have done a great job here. Now it is time for the next generation.

IV. SUMMARY ASSESSMENT AND IMPLICATIONS FOR THE NEXT STAGE

There is no doubt that the program adds value to Seattle's elementary science program, and its value potentially extends beyond that program. Below we offer a summary assessment that assumes the program should and will continue to receive support.

Summary Assessment

- The Expository Writing and Science Notebooks Program is meeting the practical instructional needs of teachers using science kits. The program thus adds considerable value to, and is an important dimension of, the kit-based science program recently implemented in Seattle schools.

- The program is contributing to student learning of science and writing across the elementary grades in ways that are consistent with the program's goals and standards, and that are also valued by the broader field.

- The program appears to be especially effective in supporting the learning of students with special needs and fostering development of students' emergent literacy. The program thus seems to have potential to contribute to more equitable outcomes for different student populations.

- The Expository Writing and Science Notebooks Program has potential for large-scale implementation because it has a well-specified curriculum, systematic professional development, and growing teacher leadership capacity.

- From the perspective of some Seattle teachers and all members of the independent panel, the program currently appears to under-emphasize uses of writing that support students' own inquiries and their pursuit of their own scientific thinking.

Implications for Further Development of the Program

We offer the following suggestions for the next phase of the program:

- Continue offering teachers across the district access to the program materials and its approach (centrally and / or at school sites) in order to grow the number of teachers involved in the program.

- The supplemental writing curriculum is currently designed so that writing tasks are intended to be optimally supportive of key science ideas in each kit, and this should

continue. Further refinement and expansion of the writing curriculum should aim at an additional purpose: offering students greater opportunity and challenge in terms of pursuing their own inquiries and scientific thinking.

- Offer participating teachers ongoing opportunities for professional development beyond the initial set of workshops. These opportunities should include joint study of student work in notebooks and shared problem-solving about classroom strategies, including adapting the writing curriculum for students with different skills.

- The program can and should offer accomplished teachers opportunities and roles for leadership that go beyond those that now exist. Key Lead Science Writing Teachers should, for example, play a major role in expanding the writing curriculum to foster student-centered inquiry, as well as facilitating ongoing professional development for participating teachers.

***

There are innumerable possibilities for further study and evaluation, both to address questions this study raises and questions it does not address. Rather than present them here, we propose that we explore with the program leaders and funders what next steps, if any, an evaluation should take. We believe it would make sense to discuss evaluation after the program's next cycle of work is fixed and after everyone involved has had an opportunity to reflect on the results of this study.