Promoting Science Learning Among English Language Learners (P-SELL)

Classroom Observation Guideline

December 2005

These scales are based on several sources, including: (a) the scales for authentic instruction and the scoring of student work developed by the Center for Organization and Restructuring of Schools at the University of Wisconsin-Madison, (b) the classroom visitation scales developed by the National Center for Research in Mathematical Sciences Education at the University of Wisconsin-Madison, and (c) classroom observation scales developed by the "Science for All" project at the University of Miami.

SECTION 1

Classroom Observation Notes

File name: _____ Observer's name: _____

School: _____ Teacher (Grade): _____

Date: _____ Time (begin/end): _____

Topic of instruction: _____

- Science concepts: _____
- Class activities: _____

Students: • Male: _____ • Female: _____ • Total: _____

General descriptions of the classroom atmosphere

Summary of Scores

Science Learning

1. Scientific understanding	1	2	3	4	5
2. Scientific inquiry	1	2	3	4	5
3. Teacher knowledge of science content	1	2	3	4	5

English Language Development

4. Teacher support of English language development	1	2	3	4	5

Student Behavior

5. Student initiative	1	2	3	4	5

Justification

1. Scientific understanding (score: XX)

2. Scientific inquiry (score: XX)

3. Teacher knowledge of science content (score: XX)

4. Teacher support of English language development (score: XX)

5. Student initiative (score: XX)

Description of the Lesson

Color-code your class description, highlighting notable sections:

pink: Scientific understanding
dark yellow: Scientific inquiry
gray: Teacher knowledge of science content
yellow: Teacher support of English language development
turquoise: Student initiative

SECTION 2

To The Observer

The purpose of these scales is to gather (high-inference) numeric indicators of the instructional environments that students experience. Teachers and their students create an instructional environment. These scales combine judgments about teachers' teaching with those about students' learning behaviors. These scales should not be thought of as passing judgment on the quality of teaching in the manner that some teacher effectiveness scales do. It is very possible for a teacher to be working very hard at achieving some goal, but students are simply not cooperating or unengaged in working with the teacher to achieve this goal. The resulting environment, where teacher and students are working at odds to one another, will be impoverished and possible not score high on some of these scales.

Your observation notes will be the primary source of *evidence* for how the classroom is scored on the scales. Hence, your notes should be detailed enough to support the ratings and descriptive enough to give another reader a good sense of what occurred during the lesson. It may help to review the scales before starting the observation and use them to help focus what you look for; that is, you are looking for evidence to support your ratings. Such evidence includes what the teacher does, exchanges involving teacher and students (the closer to verbatim, the better), counts of students who are engaged, inferences about the ambience of the class or the distribution of authority within it, documentation on how individuals or groups of students are working, copies of what is written on the board, and the like.

Remember Three Things

a. Do NOT allow an overall impression of the class to bias your ratings. The scales are intended to be analytically distinct and each scale should be applied independently of the others. Indeed, we are asking you to provide evidence to support your rating on *each* of the different scales precisely to keep these analytic distinctions.

b. All of the scales require some rather high inferences on your part. To help you calibrate your initial scoring along a specific scale, think about the spirit in which it was written. Which number best characterizes the classroom environment along this dimension? When trying to decide between two numbers, use the higher score only if you are convinced that the class characteristics meet the *minimum* criteria for that score. This does not mean that your judgment must be beyond the shadow of a doubt; you simply need to be convinced that what you observed has met those criteria.

c. To the degree possible, do the ratings soon after observing the class, before you observe another class (and wipe the impressions of the previous one out of your memory). Provide evidence for your ratings *for each scale* based on your notes – cutting and pasting sections of the narrative description into the individual justifications is a useful way to do this.

Spirit of the Scales

When applying these scales, you should think about the spirit in which they were developed. For most scales, the numbers go up based on two things: (a) the intensity or frequency with which something is taking place and (b) the number of students who are engaged in doing that thing. As a rule of thumb – and only that – it may help to think of the numbers as the following:

1. The stereotype of undesirable environment;
2. Minimal intensity; could be limited to the teacher (scales 1 and 2) or to a few students (scales 1, 2, and 5);
3. Greater and/or uneven intensity; includes some students (scales 1, 2, and 5);
4. Substantial and intense; includes many to most students (scales 1, 2, and 5);
5. Very intense; includes most to almost all students (scales 1, 2, and 5)

Meanings of Terms

Almost all (of the time, of the students): *90% or more* (of the time, of the students)

Most (of the time, of the students): *from 50% to 90%* (of the time, of the students)

Many (times, students) and/or *much* (of the time): *more then 20% and less than 50%* (of the time, of the students)

Some (of the time, of the students): *20% or less* (of the time, of the students)

A few (times, of the students): *10% or less* (of the time, of the students)

Small Groups

Very often, observers confound the existence of small groups and small-group instruction with the intellectual environment of the classroom. How often have you heard someone say: "They were involved in small group instruction," when trying to convey some sort of reformed classroom environment? Yet it is quite possible for small group instruction to be devoid of serious content, to be dominated by one or two more capable students, or in some way to fail to live up to people's expectations on how instructional groups should function. Alternately, whole class instruction can, under some conditions, be very supportive of student learning and student engagement in high quality intellectual content. Also, the fact that students may be organized physically into small groups does not always mean that they are *working as* groups, i.e. sharing information and cooperating in the realization of some tasks.

It is important for you to find out *what* is going on within the small instructional groups. You will need to go from group to group, spend enough time in a group to get a good sense of the environment that is the context for student learning. In some cases, you may miss observing other groups because you were pressed for time. Your judgments will have to be based on the group(s) that you have observed. Our assumption is that the groups you observe are, somehow, representative of what is going on within the class as a whole. Obviously, such an assumption is

subject to error; but as long as you are relatively random in the groups you go to, we can live with that problem.

Guidelines for Classroom Observation and Construction of Fieldnotes

Observe an entire class: Ratings are distorted when only part of a class is observed. Teachers may teach a science lesson over two half-hour periods, often separated by lunch or recess. In such cases, the researcher should observe both halves of the lesson. He/she may spend the intervening time to clean up the first block of observation notes.

Keep typing/writing: As a rule, you should be typing almost all the time, since there is always some activity going on in the class. In fact, an observant fieldworker finds that classroom activity proceeds too fast to get it all down; this is to be expected and should not be a cause of frustration. If you find yourself not typing, look around at what students are doing and find something else to notate. The exception to this rule is when you are moving around the classroom to observe the work of small student groups.

Conventions of Form

When constructing your observation notes, please follow these standard conventions:

- Register as many verbatim utterances as possible. However, since it would be impossible to record every verbal exchange in verbatim, it would be often necessary to summarize the gist of the exchange.
- Exchanges among multiple students should indicate "S1, S2," etc., to show the contributions of different participating students.
- Proofread and "clean up" observation notes (e.g., spelling and punctuation) as soon as possible after the class, even if you won't get around to doing ratings until several days later.
- Most, if not all, of your justifications should include fragments of the class description as evidence supporting your rating. The justification should *not* be limited to a sentence copied from the scale itself; rather, it should describe the activity.
- Color-code your class description, highlighting notable sections (or, if you find the highlighting makes the text too hard to read, you may use text colors instead). This will not only make it easier and faster to do the justifications, but it will help later if other team members want to review all the observation notes to focus on a particular aspect of instruction. The proposed color codes are the following:

  pink: Scientific understanding

  dark yellow: Scientific inquiry

  gray: Teacher knowledge of science content

  yellow: Teacher support of English language development

  turquoise: Student initiative

- Use the following conventions for your observation notes:

  Include key transition times throughout the lesson period.

(regular font)	added descriptive information (including missing information)
[*italics*]	opinions or comments
break between lines	a change of topics or activities
T:	teacher talking
S:	student talking
SS:	multiple students responding at the same time
S1, S2, S3, etc.	exchanges among multiple students
B:	boy
G:	girl

SECTION 3

Rating Scales for Classroom Observations

Scale 1. Scientific Understanding

To what extent do students demonstrate a deep understanding of science concepts? To what extent is knowledge treated in a shallow and superficial manner?

For students, scientific knowledge is deep when they develop relatively complex understandings of the lesson's concepts. They also may produce new knowledge when they connect science concepts or topics to one another. In addition, they apply science concepts to explain natural phenomena or real world situations. Instead of being able to recite only fragmented pieces of information, students develop relatively systematic, integrated, or holistic understandings of the scientific content. Students may solve problems by applying knowledge to a variety of different situations and contexts. Students demonstrate understanding of the problematic and incomplete nature of information. Students demonstrate understanding by making reasoned and well-supported arguments.

Scientific knowledge is shallow, thin, or superficial when concepts have been taught in isolation from related ideas, personal experiences, or real world phenomena, providing students with only a surface acquaintance with their meaning. This superficiality can be due, in part, to instructional strategies, such as when teachers cover large quantities of fragmented ideas and bits of information that are unconnected to other knowledge. Evidence of shallow understanding by students exists when they do not or cannot use knowledge to make clear distinctions, build arguments, solve problems, or develop more complex understandings of other related phenomena.

In scoring this item, observers should note that depth of knowledge and understanding refers to the substantive character of the ideas that students express as they consider scientific topics. It is possible to have a lesson containing substantively important and deep knowledge, but students fail to show understanding of the complexity or the significance of the ideas. Observers' ratings should reflect the depth to which *students* pursue the content.

Rating Scale - Scientific Understanding

1. Knowledge is superficial because concepts are taught in isolation from related ideas, personal experiences, or real world phenomena. Students are mainly required to memorize information.

2. Knowledge remains superficial. Underlying or related concepts and ideas might be mentioned or covered, but only a superficial understanding of these ideas is evident.

3. Knowledge is treated unevenly during instruction; there is deep understanding of some scientific concepts and ideas, but superficial understanding of some other ideas. At least one idea is presented in depth and its significance may be grasped by some students (10%-20%), but in general the focus is not sustained.

4. Knowledge is relatively deep because the students provide information, arguments, or reasoning that demonstrates the complexity of one or more ideas. The teacher structures the lesson so that many students (20%-50%) do at least one of the following: sustain a focus on a significant topic for a period of time; demonstrate understanding of the connections between concepts, and between these and personal experiences or real world phenomena; demonstrate understanding of the problematic and incomplete nature of information; or demonstrate understanding by making reasoned and well-supported arguments.

5. Knowledge is consistently deep because the teacher successfully structures the lesson so that most students (50%-90%) do at least one of the following: sustain a focus on a significant topic for a period of time; demonstrate understanding of the connections between concepts, and between these and personal experiences or real world phenomena; demonstrate understanding of the problematic and incomplete nature of information; or demonstrate understanding by making reasoned and well-supported arguments.

Scale 2. Scientific Inquiry

To what extent do students engage in scientific inquiry?

This scale is intended to measure the extent to which students engage in scientific inquiry related to the practice of science. Scientific inquiry occurs when students conduct investigations or experiments to answer questions about the natural world. Scientific inquiry involves generating questions, designing investigations and planning procedures, carrying out investigations, analyzing and drawing conclusions, and reporting findings. Inquiry is not a linear process; instead, aspects of inquiry interact in complex ways.

Scientific inquiry is high in a lesson when students engaged in tasks such as:

- Asking questions about objects, organisms, and events in the environment that can be answered through scientific observation, data collection, and interpretation.
- Planning and designing scientific investigations, including the use of original procedures to answer questions.
- Making predictions or hypotheses about what might be found during investigations.
- Using tools and techniques to gather, analyze, and interpret data.
- Searching for patterns, making inferences, and justifying inferences with evidence (data-based).
- Using simulations and models to construct reasonable explanations.
- Using mathematics in appropriate ways as applicable to a given investigation.
- Communicating and justifying scientific procedures, investigations, and explanations.
- Engaging in scientific argumentation grounded in evaluation of data and interpretations.

A lesson is low in scientific inquiry when students' activities are limited to following a scripted set of procedures that does not require them to engage in the kinds of inquiry tasks listed above, or students do not engage in scientific inquiry.

Note: Scientific inquiry might take place almost accidentally or, seemingly, as an aside to the main flow of the lesson. For example, the teacher may ask a rhetorical question whose posing, if the question were taken seriously, would provide evidence of scientific inquiry.

Rating Scale - Scientific Inquiry

1. In no activities during the lesson do students engage in scientific inquiry.

2. Students conduct scientific inquiry within the bounds of the scripted lesson. Students primarily receive, recite, or perform routine procedures for the inquiry. Or students engage in one or two inquiry activities beyond the scripted lesson as a minor diversion.

3. There is at least one significant activity involving scientific inquiry beyond the scripted lesson in which some students (10%-20%) demonstrate engagement in scientific inquiry. Or students engage in scientific inquiry beyond the scripted lesson sporadically.

4. There is at least one major activity beyond the scripted lesson in which many students (20%-50%) engage in scientific inquiry. This activity occupies a substantial portion of the lesson.

5. Most students (50%-90%), for most of the time (50%-90%), are engaged in scientific inquiry beyond the scripted lesson.

Scale 3. Teacher Knowledge of Science Content

How accurate and comprehensive is the teacher's mastery of the science content of the lesson?

This scale indicates the extent to which the teacher has an accurate and comprehensive grasp of the science content of the lesson. While elementary grade teachers are not expected to match the degree of mastery that a scientist or other specialist would have in the field, they should possess accurate information about the topic they are teaching. Their mastery of the content should be at least slightly above that expected of students upon successful completion of the lesson. Teachers should be able to answer students' questions that go beyond the bounds of the lesson, or at least indicate to students how one might go about finding out the answer and/or what factors limit the possibilities for doing so. Of course, responding "I don't know" is preferable to proffering incorrect information, but such a response should be accompanied by suggestions (or asking students for suggestions) of how students and teacher might find out more.

A high score on this scale would be characterized by the teacher responding to students' questions with relevant information beyond that included in the lesson, enriching the lesson by providing deeper knowledge of the phenomena, or linking it to other phenomena or experiences known to students.

A low score would be characterized by multiple inaccuracies in the information that the teacher transmits to students. Students' learning difficulties or common misconceptions are provided in each lesson of the teachers' guides.

Note: This scale focuses more on teacher behavior than on students. As with all of the scales, however, the interaction between teacher and students is the focus of observation; in this case, how the teacher's mastery of the content affects the information students receive and the teacher's ability to promote students' inquiry practices. Thus, more extensive transmission of knowledge from teacher to students is not always better. The teacher's mastery of the content should not give way to long monologues that are too advanced for students to grasp, or that impede them from carrying out inquiry practices.

Rating Scale – Teacher Knowledge of Science Content

1. The teacher transmits multiple major inaccuracies (i.e., more than two) to students in his/her explanations of the phenomena under study, or makes statements that indicate a fundamental misunderstanding of the facts or processes involved.

2. The teacher transmits 1-2 major scientific inaccuracies during the lesson. His/her grasp of the science content is generally accurate, but shallow and/or tenuous. Uncertainties are not pursued with students as potential paths toward deeper understanding of the topic.

3. The teacher's knowledge appears accurate, but limited to the bounds of the lesson content. Further queries by students, if they arise, are met with responses of "I don't know" or "That's not part of the lesson," with no discussion of how one might investigate further.

4. Once or twice, the teacher transmits to students accurate and relevant information about the topic that goes beyond what is covered in the lesson. This may occur spontaneously or in response to students' questions.

5. The teacher demonstrates knowledge of the topic that goes beyond the merely adequate, enriching the discussion with "extra" information at key points in the lesson. He/she enriches the lesson by providing deeper knowledge of the phenomena. He/she is able to link the topic to other phenomena or experiences known to students in accurate and relevant ways, allowing for deeper discussion.

Scale 4. Teacher Support of English Language Development

To what extent do teachers support students' comprehension of academic content and development of English language?

Teacher support of English language development refers to how teachers adjust the level and mode of their communication (verbal, gestural, written, graphic) to enhance students' comprehension of academic content and to develop English language. With effective teacher support, teachers communicate at and slightly above students' level of linguistic competence. Teachers may also structure classroom environments in such a way as to encourage students to provide support for their peers.

Note: There may be a wide range of levels of English proficiency, as well as familiarity with scientific terminology, within a single classroom. The scale refers to the teacher's adaptation of his/her use of language to address all of these levels, not just one (be it the highest or the lowest). Language use includes both everyday communicative language and more technically oriented scientific terms, phrases, clauses, locutions, and the like.

A high score on this scale indicates that the teacher engages in several aspects of support for comprehension of academic content and English language development to move the lesson forward substantively.

First, teachers appropriately structure activities to reduce the language load required for participation (e.g., slower rate, enunciation). They use language that matches students' levels of communicative competence in length, complexity, and abstraction, such as reducing difficult language to key vocabulary or using shorter utterances and simplified sentence structures.

Second, teachers communicate at and slightly above students' levels of communication (comprehensible input). For example, during a lesson that involves the concepts of "increase" and "decrease," a teacher in a class with many emergent speakers of English helps them understand by also using the terms "go up" and "go down," as well as hand gestures or drawings. In another class, where students are more English proficient, a teacher builds on the concepts with more scientific vocabulary, such as "expand" and "contract" (linguistic amplification). In both classes, the teachers are promoting English language proficiency, while helping their students to understand scientific concepts.

Finally, teachers may also use language support strategies with ELLs, including:

- use of key terms and definitions of these terms to support comprehension and English language development,
- multiple modes of representation using non-verbal (gestural), oral, graphic, and written communication,
- use of language in multiple contexts (e.g., introduce, write, repeat, and highlight),
- use of realia (demonstration of real objects or events),
- use of ELL's home language as needed, such as using the science vocabulary terms provided in the P-SELL units, allowing ELLs to discuss in class using their home language, encouraging bilingual students to assist less English proficient

students in their home language, or allowing ELLs to write about science ideas or experiments in their home language.

A low score on this scale indicates that the teacher fails to use such techniques successfully so as to support comprehension of academic content and English language development for students. Teachers might not be aware of students' linguistic competence, or they might not know techniques for supporting students at different levels, or there might be some other reason (or combination of reasons); regardless, the teacher does not use these techniques.

For example, teachers may fail to adequately adjust their verbal communication to students' level and regularly communicate at a level beyond some students' comprehension. Conversely, teachers may consistently "lower the bar" to accommodate the least proficient students, communicating at levels that fail to challenge other students or help increase their level of competence.

Note: Some techniques may serve as appropriate support in some settings but not in others. For example, a teacher paraphrasing the same idea in different ways may help students' comprehension in some instances but confuse the students in other instances. A high score on this scale indicates that the teacher recognizes and adjusts accordingly in such situations.

Rating Scale - Teacher Support of English Language Development

1. The teacher rarely or never communicates at the appropriate level or mode of language to enhance students' comprehension and to develop English language (the level of communication is either too high or too low or is not varied to accommodate students with different levels of English proficiency). The teacher fails to use language support strategies.

2. The teacher sometimes fails to communicate at the appropriate level or mode of language to enhance students' comprehension and to develop English language. The teacher attempts language support strategies for a few students a few times.

3. The teacher communicates at the appropriate level or mode of language to enhance students' comprehension and to develop English language. There are minor events as diversions in which the teacher effectively uses language support strategies.

4. The teacher communicates at the appropriate level and mode of language to enhance students' comprehension and to develop English language. He/she uses varied language support strategies occasionally. There are one or two significant events in which the teacher effectively uses language support strategies.

5. The teacher communicates at the appropriate level and mode of language to enhance students' comprehension and to develop English language. He/she uses varied language support strategies throughout the lesson. There are several significant events in which the teacher effectively uses language support strategies.

Scale 5. Student Initiative

To what extent do students take the initiative in learning science?

This scale is designed to measure the extent to which students take the initiative in determining the direction of science-related activities and discussions. Scientific understanding and inquiry require students to go beyond simply following a procedural recipe or registering information transmitted by the teacher or text. While paying attention, listening, observing, and being engaged are important elements of science learning and not all students will be equally active in the development of a lesson, student passivity should not be the general rule. There should be evidence that students are assuming some responsibility for their independent learning, rather than being completely dependent on the teacher. Students' initiative is an indication that they are truly beginning to participate as members of a science learning community. Absence of initiative could indicate that students are disengaged from the lesson or that the teacher has near-absolute control of classroom discourse. For example, if students are not allowed to speak unless they are called upon to do so, the teacher controls the lesson and students have no opportunities to take the initiative in their own learning.

The term "student initiative" refers to unelicited comments by students that are directed at the teacher or the class in general. They do not include student responses to questions posed by the teacher. If a student raises his/her hand and is called upon by the teacher, this counts as an initiative only if the hand-raising is not solicited by the teacher. Many of students' unsolicited comments may have nothing to do with the science lesson, for example, "Miss X, Johnny wrote on my paper!" or "Can I sharpen my pencil?" For the purposes of this scale, only those comments relating to the science lesson will be considered "student initiatives." These fall into three distinct categories:

Procedural: student initiatives that refer to the procedures of the science lesson. Examples: "You have to read the graduated cylinder at eye level." "Where on the chart do I write my answer?"

Factual: student initiatives related to the factual content of the lesson. Examples: "Mine says 78 degrees." "At what temperature does water boil?" "I know the names of the nine planets."

Conceptual: student initiatives that go beyond simple factual content to clarify, elaborate, or question underlying science concepts. Examples: "If gasoline is a liquid, why is it called 'gas'?" "Air isn't matter 'cause it doesn't weigh." "When the bathroom mirror fogs up, is that condensation?" "Tornados are just like hurricanes except they're smaller."

A lesson is scored high on this scale when many students take initiatives, including some conceptual initiatives. In contrast, a lesson is scored low when student initiatives are infrequent and are limited to the procedural and factual.

Note: The issue of whether the student comment is scientifically correct or not does *not* affect its status as a student initiative. This scale measures *student* behavior; the teacher's

response to student initiatives should not directly affect the score (though it probably will affect students' willingness to initiate), but should be noted in the description of the lesson.

An accurate measure of student initiatives requires that the description of the lesson be sufficiently detailed as to capture all those that occur. If the observer is unable to register the exact content and wording of all student initiatives, he/she should capture the general content or meaning of the initiatives.

Rating Scale - Student Initiative

1. No student initiatives are observed.

2. Very few student initiatives (1-3) are observed, and these are limited to the procedural and/or factual. OR, there are more than several procedural and/or factual initiatives (3 or more), but they all come from 1 or 2 students.

3. Several student initiatives (4 ore more) are observed, and they are distributed among several students (10-20%). OR, there are very few student initiatives (1-3), but some of these are conceptual.

4. Many students (20-50%) offer initiatives, including some conceptual (1-3) ones.

5. Many students (20-50%) offer initiatives, and several of these (4 or more) are conceptual. Student initiatives play a significant role in directing class discussion and activity.