Curriculum materials. The curriculum materials used in Cycles 3-5 of the Interactions and Motion unit are provided in Appendix 1.

The CIPS Motion simulator, a java-based program that runs on Microsoft's Internet Explorer, was used for computer modeling. The software was used to model such physical phenomena as energy transfer, force, and linear motion.

Student learning. One possible concern of this study is that, even if students engaged in significant amounts of SMD, they may not have learned very much. Table 3-5 shows that most students in groups 1 and 2 did, in fact, improve their understandings of the benchmark ideas in the three curriculum cycles of interest, as measured by ranking students' initial and final understandings of these benchmark ideas as either low, medium, or high. My assessment of students' understandings were based on their Power Drive ("test") scores, their written work, and their statements during group discussion.

Table 3-5. Students' initial and final understandings of the unit 2, cycle 3-5 benchmark ideas.

		Cycle 3		**Cycle 4**		**Cycle 5**	
		Initial	***Final***	***Initial***	***Final***	***Initial***	***Final***
Group 1	Darla	Medium	High	Medium	High	Low	High
	Lacey	Medium	High	Medium	Medium/ High	Low	Medium/ High
	Porter	?	?	?	X	X	X
	Grace	Low	Low	Low	Low	Low	Low
Group 2	Roxanne	Medium	High	Low/ Medium	Medium/ High	Low	Medium
	Arthur	Medium	High	Medium	High	Low	High
	Sabrina	Medium	Medium/ High	Low/ Medium	Medium/ High	Low	Medium
	Jasper	Low	Low	X	X	X	X

Note. An **X** indicates that the student was no longer in the study at that time. Porter's initial and final understandings can't be assessed accurately because, when he returned to homeschooling, all of his written materials were inadvertently thrown out by another student.

Sample end-of-cycle Power Drive assessments for each student can be found in Appendix 2.

Data Collection

To document students' small-group discussions, their associated written work, and classroom instructions and interactions, the following data was collected:

- Daily videotapes of the entire classroom
- Daily whole-period videotapes of the two selected student groups
- Daily ethnographic fieldnotes meant to supplement the daily classroom and group videotapes
- Students' written work in their CIPS notebooks
- Photocopies of student assessments (Power Drives)

- Photocopies of student homework

Data Analysis

Overview

To document the two groups' sense-making discussion (SMD) and identify those factors supporting or hindering SMD, the following procedure was followed.

1. Identify those sub-sections of each activity where, based on the structure of the CIPS pedagogy, students are expected to engage in significant SMD. (Henceforth, I will refer to these sub-sections as "relevant sub-sections".)
2. Document instances of sense-making in relevant sub-sections according to the six-component scheme (described in Data Analysis Step 2, below)
3. Determine the distribution of sense-making instances for each group, student, activity, activity sub-section, and curriculum cycle
4. Calculate the average percentage of time dedicated to sense-making discussion for each group, activity, activity sub-section, sub-section type,[2] and curriculum cycle
5. Identify significant differences in sense-making across groups, students, activities, activity sub-sections, sub-section types, and curriculum cycles

[2] By sub-section "type", I mean one the following reoccurring sub-sections: "Prepare your wipe board!", "What does your team think?", "What really happens?", "Making sense", and "What are your team's ideas now?"

6. Determine the factors most likely to explain the differences found in step 5

Each of these steps is described in detail below.

Rationale

The primary reason for using the method outlined above is that the method closely parallels classic scientific methodology: building an expectation, getting a result, and explaining that result.

In this study, expectations for small-group SMD were based on identifying those portions of each activity where it was presumed that students would engage in scientific discussion with their group members. Focusing only on these relevant sections, the hope would then be that each activity, activity sub-section, and curriculum cycle would provoke the same amount of sense-making as every other activity, activity sub-section, and curriculum cycle, respectively. It was also the hope that each group and student would engage in sense-making discussion as much as each other group and student, respectively.

The "results" were the quantitative summaries of the number of instances of sense-making and percentage of time engaged in SMD. The "explanation" step of this method then consisted of the documentation of those factors that contributed to differences in instances of sense-making or percentages of time engaged in SMD across groups, individual students, activities, activity sub-sections, or curriculum cycles.

Limiting my investigations to those portions of the curriculum where SMD had been expected was extremely practical in that, while the appearance of SMD in places where it is unexpected is an interesting result, curriculum designers and teachers would be most interested in modifying curriculum areas where SMD is expected but does not appear.

Also, as laid out, all but the first step of the analysis are directly related to one of the three research questions. Steps 2-4 help to answer the second research question: To what extent do students engage in SMD? Steps 5 and 6 provide the method of answering the final research question: Which factors provide support for students' SMD?

Data Analysis, Step 1: Identifying Curriculum Areas where SMD is Expected

The sub-sections found in each activity from Unit 2, Cycles 3-5 are listed in Appendix 3.

Of these activities and their component sub-sections, it was expected that students would engage in significant SMD in sub-sections meeting the following criteria:

- at least one question in the sub-section asked students to explain a concept, provide evidence, explain their reasoning, predict an experimental result, or construct a graphical representation
- the sub-section was not labeled as one that was to be filled out individually
- the sub-section did not consist solely of teacher-led demonstrations or whole-class discussions

The sub-sections meeting these criteria have been italicized and bolded in Appendix 3.

The inclusion of the first criterion ensured that I could limit my analysis to those portions of the activity having the greatest potential for drawing out small-group SMD. Since underlying explanations and predictions are explicitly defined as components of SMD (see below), one could reasonably hope that sub-sections requiring students to think about and write down their explanations and predictions would provoke group discussion about those same explanations and predictions (i.e., would provoke SMD). Similarly, one could reasonably hope that sub-sections requiring students to think about and write down their scientific evidence and reasoning would provoke the sorts of concept-, inference-, and explanation-related discussions characteristic of SMD.

Examples of questions/directives from sub-sections meeting the first criterion are:

> "Can pulls be combined? (Write your team response and evidence.)" [from "Prepare your team wipe board", cycle 3, activity 1]
>
> "How does the surface roughness affect how hard it was to pull one surface across another?" [from "Making sense", cycle 4, activity 2]
>
> "Do you think that the earth's spinning causes gravity? What is your evidence? Discuss this with your team, then write your answer here." [from "Is gravity caused by the earth's rotation?", cycle 5, activity 1]

As illustrated, requests for explanations, predictions, etc. were not required to contain those exact words. The second example does not contain the word "explain", for instance, even though the question asks for an explanation. It also was not a requirement for questions/directives to explicitly ask the students to discuss something with their group members. (This is done in the third example, above.) This is because it was expected that students would engage in group discussion without always having to be told to do so.

Regarding the other criteria for sub-sections being acceptable for analysis, an example of a sub-section failing the second criterion is "Making sense" in activity 4, cycle 5. Although it was similar to identically-named sub-sections in other activities, in this instance it was labeled with the phrase "Do this part by yourself!". Examples of sub-sections failing the third criterion were "Computer simulations" in activity 1, cycle 4 (teacher-led demonstration) and the "What ideas does your academy have?" sub-sections (whole-class discussions) found in each of the Unit 's consensus activities.

Data Analysis, Step 2: Documenting Instances of SMD

In this step, students' small-group sense-making discussion was initially analyzed by identifying the instances where group members engaged in the six components of SMD. These six components are as follows:

- Predicting a phenomenon or experimental outcome
- Clarifying the facts of a phenomenon or experimental result
- Describing and explaining a phenomenon or experimental result
- Defining, describing, and connecting scientific concepts, procedures, processes, and representations
- Testing knowledge compatibility

- Making a request for any of the above

The instances of "Making a request" were broken down into each sub-type: requesting a prediction, requesting an underlying explanation, requesting a test of knowledge compatibility, and so forth.

Note that my six components of sense-making discussion (SMD) are quite different from Hatano's four original components of nonverbal sense-making (i.e., comprehension activity). A detailed description of the six components, which includes an account of how the components evolved from Hatano's four original components, can be found in Chapter 4.

General inclusions and exclusions in the identification of SMD. It was not the case that every question or comment in students' small-group discussion was a possible instance of SMD. Much of what was said in the two groups, for instance, revolved around negotiations of actions, materials, or status -- not the verbal negotiation of scientific meaning that I would call "sense-making discussion". Typically, the quickest and easiest way of determining whether a comment should be counted as SMD was whether it could instead fit into one of the other non-SMD categories of negotiation: actions, materials, or status.

Likewise, some small-group conversation was directed at making sense of something -- but not necessarily directed at the subjects that were of interest in this study. For example, students sometimes spent time trying to figure out an experimental apparatus (How does this thermometer work?) or written instructions (What are we supposed to do here?). While these exchanges are clearly examples of sense-making discussion, they are only relevant to this investigation to the extent that they detract from the discussions of scientific concepts, scientific processes, and physical phenomena. (In other words, the exchanges gain significance in Step 6 of the analysis, which is the portion dealing with the factors affecting SMD.) For the purposes of this study, SMD was only labeled as such when it was directed at understanding the Force and Motion unit's benchmark ideas, or directed at understanding experimental results, physical phenomena, and scientific procedures, processes, or representations directly related to those ideas.

Other types of statements that were excluded from consideration were yes/no answers and anything read directly from a piece of paper or wipe board. The yes/no answers were excluded because of the difficulty in discriminating between "thoughtful" yes/no answers and guess-like yes/no answers. The reading of written material was excluded because of the fact that reading something aloud is something that is typically done before or after one constructs a new understanding -- not during the construction itself. On the other hand, statements were considered to be candidates for SMD if they were uttered and written down simultaneously. This is because the student's reading, writing, and construction of meaning were inseparably intertwined during this time.

Finally, any SMD that arose in discussions with the teacher were excluded from analysis. This is because the study focused only on intra-group discussion, not the discussion that occurred between group members and people outside the group. However, sense-making discussions between the group and teacher were considered in Step 6 of the analysis. Why? Because teacher/group conversations are potential factors affecting intra-group SMD in that teacher/group conversations can have a lingering effect on conversations that take place after the teacher leaves the group.

Data Analysis, Step 3: Determining the Distribution of Sense-making Instances

As group conversation ebbs and flows over the course of an activity, some sense-making statements are directly prompted by the curriculum materials, others respond to (build on, connect to) previous sense-making statements, and still others are isolated outbursts with little connection to anything read or uttered previously. Over time, one would expect these various statements would form a distribution: a certain percentage would be clarifications, a certain percentage would be tests of knowledge compatibility, and so forth. These distributions are of interest because they allow researchers to form an idea of which components of sense-making are most and least commonplace, and also how these distributions change over groups, activities, and other factors of interest.

A number of sense-making distributions were constructed for this study. These distributions consisted of the number of times that each component of sense-making discussion (clarification, prediction, etc.) appeared in student small-group discussions. An overall distribution of all instances of sense-making was constructed, as were distributions for each group, student, activity, activity sub-section, and curriculum cycle.

Data Analysis, Step 4: Calculate the Average Percentage of Time Dedicated to Sense-making Discussion for each Group, Activity Sub-Section, Activity, and Curriculum Cycle

When participating in group activities, groups engage in a number of different activities, all of which are important. These activities include writing answers, setting up of experimental apparatus, recording data, and engaging in various types of discussion: negotiations of meaning, actions, status, materials, and also off-task discussion. Calculating the percentage of time that groups engage in SMD provides a measure as to how much time each group spends (relative to all of these other activities) on the discussion of meaning -- discussions which, ideally, are key in CIPS in helping students understand and develop ideas about how the physical world works.

The following formula was used to calculate percentage of total time spent on sense-making discussion by a particular group in a particular sub-section: (Total time the group spent on sense-making discussion) ÷ (Total time that the group spent on the sub-section).

The numerator (total time on sense-making discussion) was calculated by summing all those times where at least one member of the group was engaged in verbal sense-making. The denominator (total time in sub-section) was calculated by subtracting the time code corresponding to the start of the sub-section (i.e., when first group member started the sub-section) from the time code corresponding to the end of the sub-section (i.e., when the last group member finished the sub-section or ran out of time, whichever came first). There was only one possible adjustment to this time: if the teacher stepped in to engage the class or group in conversation/lecture, the time taken up by the teacher was subtracted from the total time; this is because there was little to no opportunity for group discussion during this time. Note: if the teacher gave the groups a certain amount of time to finish a sub-section, and the group finished early, only the time that was actually spent on the activity was counted (i.e., the "unused" portion of the time was not included in the total time).

After each group's percentage of time sense-making was calculated for each relevant sub-section in the curriculum, the average values were calculated for each group, activity sub-section, activity, and curriculum cycle. The method of calculating each of these averages is outlined below.

<u>Activity sub-section</u>. The percentages for the two groups were averaged together.

For instance, the percentages for groups 1 and 2 in "What really happens?" in activity 2 of cycle 3 were 2% and 6%, respectively. The average percentage of time engaged in SMD for this sub-section was therefore 4% [(2% + 6%) ÷ 2].

In sub-sections where one or both of the groups failed to spend any time, no average value was calculated. (This occurred in twelve sub-sections.) Groups failed to spend time on a sub-section for one of two reasons: 1) the entire group was absent, or 2) the group simply skipped the sub-section (either on its own -- whether accidentally or intentionally -- or as a result of the teacher's actions).

Activity. Once the average percentage was calculated for each individual sub-section in an activity, the sub-section averages were themselves averaged together to obtain an overall average for the activity as a whole.[3]

For example, activity 1 in cycle 3 consisted of three subsections: "What does your team think?", "What really happens?", and "Prepare your wipe board!" The average percentage of time engaged in SMD for each sub-section (unrounded) was 37.5%, 6.5%, and 8%, respectively. The overall average for activity 1 in cycle 3 was therefore 17% [(37.5% + 6.5% + 8%) ÷ 3, rounded to the nearest percentage point].

However, a problem arose in activities that contained a sub-section where one or both groups failed to spend any time (for the reasons described above). In cases where both groups failed to spend time in a sub-section, the sub-section was omitted from the calculation. (This occurred four times). In cases where only one group failed to spend time in a sub-section, the group's nonexistent percentage was replaced with a projected percentage for that group. (This occurred eight times.) The projected percentage was determined by calculating a reasonable estimate of the percentage that the group probably would have spent on SMD had it not skipped the section. This projected percentage was then used to calculate a projected average for the sub-section, which in turn was used in the overall average over all sub-sections.

[3] This was done to ensure that each sub-section is given equal weight in an activity. It is not desirable to give more weight to an activity sub-section just because it is longer or shorter than the other sub-sections in the activity.

The projected percentage for a skipped sub-section (with one skipping group and one nonskipping group) was calculated as follows: (the percentage of time that the nonskipping group spent engaged in SMD for that sub-section) x (the ratio of the skipping group's sense-making percentage for the entire cycle to the nonskipping group's sense-making percentage for that cycle). A sample calculation: Group 1 was engaged in SMD 38% of the time in the "What does your team think?" sub-section in activity 2 of cycle 3, while group 2 skipped the section. Group 2's nonexistent percentage was replaced with a projected percentage of 15%, which was calculated by multiplying 38% (group 1's percentage in the sub-section) by .39 -- which was the ratio of group 2's overall sense-making percentage for cycle 3 (9%) to group 1's overall percentage (23%). In other words, since group 2's sense-making percentage in cycle 3 was typically .39 of group 1's percentage, group 2's nonexistent percentage was replaced with 38% x .39, or 15%.

The replacement of nonexistent percentages with projected percentages was performed for activity-level averages to ensure a fair comparison between activities,[4] the justification being that -- if a group's extremely high or extremely low percentage was taken as being representative of that sub-section -- that sub-section's artificially inflated (or deflated) percentage could drastically affect the overall percentage for the activity as a whole.

---

[4] This also ensured a fair comparison between cycles, since the average percentage of time engaged in SMD in a given cycle is based on the activity/sub-section percentages in that cycle.

<u>Group</u>. The overall average for a particular group was calculated by first determining the group's average percentage in each activity. The activity percentages in a given cycle were then averaged together to attain a cycle average (for that group) -- with the important exception that "Putting it all together activities" were excluded from this analysis.[5] Finally, the cycle percentages were averaged together to provide an overall average for the entire three-cycle curriculum.

For example, group 1's activity averages in cycle 3 (omitting the final consensus activity) were 23% for activity 1, 27% for activity 2, and 40% for activity 3. Group 1's average for cycle 3 was therefore 30% [(23% + 27% + 40%) ÷ 3]. Using a similar method of calculation, group 1's averages for cycles 4 and 5 were determined to be 24% and 25%, respectively. Consequently, Group 1's overall average for the percentage of time engaged in sense-making discussion was calculated to be 26% [(30% + 24% + 25%) ÷ 3].

<u>Sub-section type</u>. For each cycle, percentages from sub-sections of a particular type (e.g., "What does your team think?") were averaged together to form a cycle average for that sub-section type. The three cycle averages for that type were then averaged together to form an overall percentage. Sub-sections of the given type were not included in the calculation if they

---

[5] These consensus activities were not included when calculating the overall percentages for each group because the opportunities for SMD in the consensus activities arose from connecting and reconsidering ideas from prior activities, the result of which was that groups who had engaged in significant SMD in these prior activities often had little need for SMD during the consensus activity. Therefore, the percentage of time engaged in SMD during the consensus activity was omitted from the overall calculations to ensure a fair comparison between groups.

were skipped by both groups; if only one group skipped the sub-section, a projected percentage was substituted for the nonexistent percentage (as described above).

For instance, in cycle 3, groups spent 37.5% of their time on SMD in activity 1's "What does your team think?" and 26.5% of their time on SMD in activity 2's "What does your team think?", giving rise to an overall average of 32% [(37.5% + 26.5%) ÷ 2] for "What does your team think?" sub-sections for the cycle. "What does your team think?" cycle averages for cycles 3 and 4 were 25% and 19%, respectively, and so the overall average percentage for "What does your team think?" subsections was calculated to be 25% [(32% + 25% + 19%) ÷ 3].

Note that many of the "What really happens?" sub-sections (i.e., sub-sections involving hands-on experiments) were not explicitly labeled as such, even though they were counted as being of this sub-section type.

<u>Curriculum cycle</u>. For each cycle, activity SMD percentages were averaged together to form a cycle average. For example, activities 1-5 in cycle 5 had percentages of 9%, 14%, 39%, 22%, and 1%, respectively. Consequently, the overall percentage for the cycle was calculated to be 17% [(9% + 14% + 39% + 22% + 1%) ÷ 5].

Data Analysis, Step 5: Identifying Significant Differences in Sense-Making across Groups, Students, Activities, Activity Sub-Sections, and Curriculum Cycles

To identify patterns of sense-making that needed to be explained in terms of factors, the following questions were posed:

1. Was the overall distribution of sense-making instances surprising in any way?
2. How did the distribution of sense-making instances (in terms of the six components of SMD) differ across sub-sections, sub-section types, activities, groups, and curriculum cycles?
3. How did the percentage of time spent on sense-making differ across sub-sections, sub-section types, activities, groups, and curriculum cycles?

An integral part of answering the first question (about the overall distribution of instances) was to set up an expectation for what the overall distribution might look like. My assumption, which was to be tested empirically, was that the vast majority of sense-making discussion would be driven by the written curriculum. If this were true, it would be reasonable as a first-order approximation to assume that the overall distribution of verbal sense-making instances should mirror the number of times that students were asked in the materials to engage in the various components of SMD. For example, if 3% of questions in the curriculum materials ask groups to make a prediction, one might assume that predictions would make up roughly 3% of the overall distribution of verbal sense-making instances.

Since my assumption about the curriculum driving the groups' SMD turned out to be true, the overall distribution of sense-making instances was compared to the distribution of sense-making components as represented in questions, directives, and graphing/diagramming activities in cycles 3-5 of the Force and Motion unit:

Table 3-6. Distribution of sense-making components in the curriculum materials.

	CL	**P**	**UE**	**DDC**	**TC**
Curriculum materials	25 (20%)	18 (14%)	26 (21%)	52 (41%)	5 (4%)

Note. Top values are actual numbers. Bottom values (in parentheses) are percentages. The components of SMD are abbreviated as follows: **CL** (**Cl**arifying the facts of a phenomenon or result), **P** (**P**redicting), **UE** (Describing or providing an **u**nderlying **e**xplanation for a phenomenon or result), **DDC** (**d**efining, **d**escribing, clarifying, and **c**onnecting scientific concepts, procedures, processes, and representations), and **TC** (**t**esting knowledge **c**ompatibility).

Answering questions 2 and 3 (relative comparisons between distributions and percentages of time engaged in SMD) involved simple numerical comparisons across sub-sections, sub-section types, activities, groups, and curriculum cycles

In answering questions 1-3, differences and variations were unearthed in the sense-making data -- the most prominent of which needed explaining in terms of personal, group, task and contextual factors that contributed to why these differences and variations existed. For example, once it was established that there was a marked difference between the two groups in terms of their percentage of time sense-making, I identified and verified the factors that made the greatest contribution to this difference between the groups.

A word on reliability. Utmost care was taken to ensure that the sorting of students' sense-making statements was done consistently. In fact, as my six components of sense-making were still being developed, my inability to decide whether a statement should fall into one of two categories was often a tip-off that a component of SMD needed to be slightly redefined, or perhaps even that a new component of SMD needed to be created. Through the process of identifying and resolving inconsistencies, my components (categories) were refined until sense-making statements could be sorted with confidence.

In calculating the percentage of time that a group spent on SMD, I was very conservative in classifying discussions as sense-making or not sense-making. If I wasn't sure that something was genuine sense-making discussion, I classified it as non-SMD.

Despite my care in sorting statements and identifying SMD, there were bound to be a handful of codings that were inconsistent. However, it is important to point out that the major focus of this study was not on sense-making statements or percentages of time spend on SMD, but instead on significant differences across students, groups, and curriculum areas on these same variables. So if a small number of clarifications were mis-classified as underlying explanations, or if I was off a percentage point or two for the amount of time that a group spend on SMD, these errors would not affect the results of interest: that there were far more clarifying statements than expected, for instance, or that group 1 had a much higher sense-making percentage than group 2. In other words, a small number of coding errors would not be enough to affect the significant differences in sense-making across students, groups, and portions of the curriculum (cycles, activities, etc.), which were the actual results of interest in this study.

Data Analysis, Step 6: Identifying and Verifying the Factors Supporting (or not) Small-Group SMD

Once I identified differences in SMD between groups, activities, activity sub-sections, and cycles, I began the process of identifying factors (or interaction between factors) that might help explain these differences. This process consisted of four analytical passes through the data (transcripts and videotape):

First pass: Keeping one's mind as open as possible, identify factors that might explain prominent differences in sense-making (across groups, cycles, etc.). The only preconceived notion at this step is that any given factor will probably fall into one of four categories: activity-, group-, personal-, or context-based factors.

Second pass: Consult a preconstructed list of possible factors (which had been formulated by reviewing past research) to determine if any additional activity-, group-, personal-, or context-based factors might help explain the relevant differences in sense-making.

Third pass: Identify additional factors by looking for deviations in established patterns of sense-making.

Fourth pass: Test the validity of the list of factors by investigating other factors that might explain relevant differences in sense-making.

Each of these analytical passes through the data is described in more detail below.

First pass: As a first step, I re-read the transcripts and re-watched the videotapes, during which time I recorded activity-, group-, personal-, and context-based factors that appeared to have a significant effect on the result of interest. As this was done, time codes from the transcripts/videos were recorded to mark particular statements, conversations, or actions that demonstrated the effect of these factors.

Second pass: Next, I consulted my "possible factors" list to help me identify additional factors that may have played a role. This list had been formulated by summarizing the results of my literature review (chapter 2) on factors having an effect (or probable effect) on sense-making discussion. The list:

Table 2-3. Factors having a possible effect on the sense-making discussion in this study. (continued)

Personal factors		**Group factors**	
Student goals	Prior knowledge	Group expectations	Shared goals
Subject matter interest	Freedom from need	Leadership style	Collaborative skills
Cognitive incongruity	Metacognition	Student roles	
Comprehension: important and possible?			
Contextual factors		**Task factors**	
Classroom expectations	Role of teacher	Task goals	Science content
Assessments		Materials	Intrinsic motivation

If group members' statements, conversations, or actions demonstrated the effect of a new factor from the list, the new factor and supporting timecodes were added to the result's factor list.

In these first two passes through the data (one pass without the list, one pass with the list), establishing a factor's effect on SMD was often straightforward, due to the fact that certain factors had clear and direct effects on students' statements, actions, or facial expressions. For example, students' exclamations of "Duh! This is too easy!" or "We all knew how the experiment would turn out!" were clear indicators that an activity was pitched at a level of understanding that was too low for the group, and that therefore the students felt little need to discuss the ideas or experimental results associated with that activity. Another example of a clear factor was that, when Sabrina (group 2) turned to Arthur to help her clear up her conceptual confusion, Arthur sometimes responded with a belittling and condescending tone -- the result of which was that Sabrina would discontinue the conversation, and so group 2's sense-making discussion would be prematurely cut short.

<u>Third pass</u>. Factors whose effects were not obvious from the first two passes through the data were identified in the following manner. These factors (and their effects) were investigated by examining deviations in established patterns of sense-making. For example, Darla's statements and actions (in group 1) often demonstrated a general interest in science and a general concern for the learning of her groupmates -- both of which were stable personal factors that helped support her group's SMD. Therefore,

when Darla did not exhibit this interest or concern, it was assumed that there were other personal, group, task, or contextual factors that significantly influenced the group's SMD (i.e., suppressed Darla's interest and concern) during this period. Likewise, when Grace engaged in significant SMD (which was very unusual for her), it was assumed that there was a factor specific to the situation or materials that helped support her sense-making.

Fourth pass. Once I had decided on an initial list of factors that might explain an interesting result, the next step was to investigate whether other factors might explain the effects attributed to each factor. In other words, the next step was to answer this question: Is my list of factors valid, or are other factors actually causing these effects? An important aspect of this portion of the analysis was in determining whether certain factors were stable over time, groups, activities, activity sub-sections, or cycles. For instance, if a student only demonstrated significant interest in four activities out of twenty, then it could not rightfully be claimed that the student had a general interest in science, and that it was this general interest that contributed significantly to the differences between the groups' SMD.

For each factor, as in the example above, a test of validity was performed by examining the factor's supposed effect on student behavior and student thinking (as indicated by student statements, actions, etc.) and brainstorming other factors that might also reasonably lead to the same effect. I then looked for evidence for the existence of these other factors. If no evidence could be found, then the original factor was kept on the list as a valid factor, and the other factors were dismissed as possible candidates (in this case only, as they may be relevant to other differences in sense-making). If significant evidence existed that pointed to the existence of these other factors -- and also pointed to the nonexistence of the original factor -- then the new factor was added to the list of factors for that result, and the original factor was discarded. If significant evidence existed that supported the existence of both factors (the original and the new factor), then the new factor was added to the list and a note was made that both factors likely shared some responsibility for the effects seen in the transcripts/video.

After the test of validity -- in which new factors were added to or subtracted from the list, or the effects of the original factors were modified -- the "final" list of factors for each result was used to explain the differences between sub-sections, sub-section types, activities, groups, and curriculum cycles in the quantitative measures of sense-making: the number of instances of sense-making and the percentage of time engaged in sense-making.