Any episode of learning and teaching is necessarily situated in both space and time. But, whereas the spatial arrangement of the classroom remains relatively constant, change is the very essence of the learning that takes place within it. Such is the nature of the data to be examined in this paper. The same curriculum unit was taught by the same teacher to a fourth-grade class in four successive years and every lesson in each year was videorecorded. Our purpose was to document change over time at three levels: from year to year, within each year, and over the course of activities within particular lessons. Within the limits of this paper, we focus on one particular issue as it recurred over time.

Keywords: dialog; inquiry; knowledge building; social constructivism

Exploratory and explanatory talk

Over the last half century there has been growing acceptance that learning is an active and social process. The constructive nature of learning was enunciated by Piaget (1970) and has since been accepted by the majority of educational researchers (Case 1996). To this, Vygotsky (1978) added an emphasis on the social nature of learning, both in relation to the provenance of what is learned and to the processes through which learning takes place. As is well known, Vygotsky (1987) proposed that learning is most effective when the learner receives assistance in her or his zone of proximal development while engaged in some purposeful activity with an adult or more capable peer. In this process, talk is a major means of mediation together with collaborative action. And, as Bakhtin, Vygotsky’s contemporary, made clear, the talk needs to be dialogic, in the sense that “the speaker does not expect passive understanding that, so to speak, only duplicates his own idea in someone else’s mind. Rather he expects response, agreement, sympathy, objection, execution, and so forth’ (1986, 69). In other words, learning involves the coconstruction of knowledge rather than its one-way transmission from expert to novice.

Studies of language development in the early years have provided strong support for the key role of caretakers as dialogic partners in children’s learning (Bates 1976; Bruner 1983; Nelson 1996; Nelson 2007; Wells 1986) and similar evidence has been advanced for the role of teachers in promoting effective language learning and learning through language during school years (Alexander 2006; Mercer 1995; Nystrand 1997; Wells 1999). As these social scientists argue, students’ understanding of new material is most effectively achieved when curriculum is coconstructed through dialog, in which students are able to make connections...
T. Ball and G. Wells

Figure 1. The spiral of knowing (adapted from Wells 1999).

to previous learning both in and out of school and are encouraged to voice their ideas and opinions in the knowledge that they will be taken seriously by the teacher as well as their peers.

This social constructivist conception of learning is captured in what Wells (1999) called the spiral of knowing (Figure 1). Each cycle through the spiral starts from what the knower brings to the new situation from the sense that she/he has made of previous experience and it ends, ideally, in enhanced understanding of the information that is made available through feedback from action or from other participants in the situation (teacher, text, or other more expert source). However, for new information to lead to enhanced understanding, the learner needs to engage in some form of knowledge building through dialog with other participants in the situation and/or through exploratory or hypothesis-testing action with respect to an object that the learner constructs and/or attempts to improve.

The importance of opportunities for students to engage in such knowledge building through discussion was demonstrated many years ago by Barnes (1976) and Barnes and Todd (1977), and has since been supported by a considerable body of research across the learning sciences (Scardamalia and Bereiter 1991, 2006) and especially by studies in science education (Driver 1983). Central to knowledge building, it has been argued, is the making, critiquing, and defending of explanations about the topics and issues which students are exploring. Furthermore, following Vygotsky’s argument that ‘higher mental functions’ develop through the appropriation of interpersonal dialog, participation in collaborative knowledge building enables learners to develop the ability to engage in comparable individual mental activity in the medium of inner speech.

This line of thinking has given rise to a number of intervention projects designed to develop students’ mastery of collaborative knowledge building through participation in what Mercer (2002) calls ‘exploratory’ talk in the ‘zone of intermental development’. To this end, Wegerif, Mercer, and Dawes compared the performances of students who received training in ‘ground rules’ designed explicitly to foster the sort of talk, ‘in which reasoning is made visible and publicly accountable through the discussion of alternatives’ (1999, 497), with the performances of students who did not participate in such training.

By evaluating transcripts created from video recordings of focus groups discussing the problems during group exercises, the authors were able to identify key words and phrases that could be used as markers of sequences of exploratory talk, distinguishing this type of
talk from two alternative patterns: ‘disputational’ and ‘cumulative’ talk. They also found that the individual scores of students who used more exploratory talk to think aloud about reasoning problems on the Raven’s Progressive Matrices test improved significantly between pre- and post-intervention assessments, while the individual scores of students who did not participate in the intervention did not change significantly, even though untrained students were given equal time to discuss the problems in small groups.

In a more recent study by Herrenkohl and Guerra (1998), students were assigned to different ‘intellectual roles’ involved in constructing scientific explanations such as ‘making a prediction’, ‘summarizing results’, or ‘relating evidence to predictions’. Meanwhile, peer counterparts were assigned to matching ‘audience roles’, and given the responsibility for generating (or borrowing from a list of suggestions) pointed questions that would help explainers articulate summaries, predictions, and arguments more clearly. Herrenkohl and Guerra found that students participating in classrooms where explaining practices were structured through this kind of complementary role-play were more likely than students in a control classroom to discuss observations, identify mistakes, challenge claims, examine inconsistencies in their thinking, and articulate and revise their theories while making sense of observed outcomes. Moreover, Herrenkohl and Guerra (1998) provided evidence that students working in this participant structure learned to self-monitor and take more responsibility for their own understanding. In sum, the findings from these studies show that when dialogic-explaining practices are explicitly supported as a central and valued form of participation, student engagement increases as mental functions, such as reading comprehension and scientific thinking skills, improve.

In seeking to explain why participating in explanatory dialog is beneficial for students’ intellectual development, we have argued that it is particularly in the process of generating a response for others that new networks of ideas begin to cohere and individuals become aware of the contradictions or discontinuities in their own understanding (Wells 1999). Similar arguments have been put forward in the more specific field of science education where the development and articulation of scientific explanations are central and valued practices. Keil (2006) discusses how explanation and argumentation aimed at increased understanding are different from mental models, intuitive theories, and procedural knowledge. Keil suggests that explanations are operative in expanding understanding because producing explanations involves interpretive moves that create conceptual ‘trajectories’, where models and theories can remain relatively static and isolated. Emphasizing the ‘transactional nature’ (2006, 229) of explanations, Keil argues that explanation is ‘inevitably rhetorical’ and that rhetoric or ‘the effort to organize evidence (inscriptions) and claims into persuasive accounts, is a central aspect of scientific argumentation as it is actually practiced’ (p. 226). Furthermore, students who conflate theories and evidence can benefit from constructing explanations, which require students to focus on the relevance of the perceived relations connecting claims to evidence.

Finally, engaging students in exploratory and explanatory talk also serves an important pedagogical function. In order to help students build new understanding, it is important for teachers to first discover what their students already understand about domain content or relevant practices. If students are not actively offering original contributions, teachers will end up doing much more guesswork to discern student comprehension. If instead, teachers are able to draw out not only the intuitive theories or particular conclusions their students are using but also learn how they were generated, teachers will have more opportunities to employ tailored facilitation moves that support emerging understanding more strategically.
Cars on ramps

The project we are reporting here had a serendipitous beginning. On the very last day of school before the winter holiday, the second author accepted an unexpected invitation to visit a fourth-grade classroom to witness the annual ritual of the Lunch-box Derby. Students had made ‘cars’ of fruits and vegetables held together with wooden skewers and competed to see whose car would travel farthest when launched down an inclined ramp. Every student was allowed two trials and the teacher recorded the results of each trial. After the holiday, the second author approached the teacher with the suggestion that the Lunch-box Derby be treated as the ‘launch’ for a more extended investigation of the factors that affected how far a more durable vehicle could travel when launched down a ramp. The teacher agreed and, a week or two later, the students, working in groups, began to construct vehicles from scrap materials and to make modifications to them in the light of the information gained from trials on the ramp.

Our aim was to investigate the effectiveness of the unit in engaging the students in understanding the physics of motion while mastering the practical skills involved in the inquiry, and to discover whether and how this approach could promote explanatory talk. At the macro level, our purpose was to study whether and how the teacher changed in his use and management of whole-class discussion (a) as he became more experienced with the unit, and (b) as a result of the changes that took place with respect to materials and demographics. At a second level, we wanted to discover whether change occurred over the course of each year’s unit in (a) the students’ understanding of the scientific concepts involved, and (b) in their willingness and ability to explain the relationship among the changes they made to the vehicles, their measurements of distance traveled, and the forces responsible for the observed results of their actions. Finally, we planned to conduct microanalyses of the move-by-move patterns of teacher–student interaction at the level of individual sequences in order to establish whether there are particular types of initiation and follow-up that foster students’ production of progressively more complex explanations.

Research site and participants

This collaborative action research project took place in Shoreline Public School (a pseudonym) in a small city in the Central Coast region of California. The neighborhood in which the school is situated is mixed both in terms of the socioeconomic status of the families and their ethnicity, with a substantial proportion being of Mexican origin. In the first two years of the project, approximately three-quarters of the students in the fourth-grade class in which the research was carried out were of European origin. This ratio changed in year three as the result of the closure of a nearby school and transfer of the majority of its students to Shoreline. This increased the proportion of Latino/a and Mexican-origin students in the grade four class to more than half. Many of these students were classified as English language learners (ELL) but their proficiency in English was sufficient for them to be able to participate in the inquiry unit under investigation.

Buzz, the teacher of the class, is a Caucasian male with many years of teaching experience at the elementary level. His class is well known for the imaginative and innovative ways in which he designs the curriculum and for the plays that he writes and directs for his students to perform for the rest of the school, parents, and other visitors. Students in his class frequently engage in group work and generally perform well on the state’s standardized tests.

The project was undertaken by Buzz and the two authors as collaborative action research. All three of us shared in planning and teaching the lessons each year, with Buzz playing the major role.
Structure of unit and constituent lessons

Each year, the unit had a similar overall structure. Students first built their own individual vehicles at home, often with parental help, and the unit was launched with a competition in which each student released her or his vehicle from the top of the ramp and measured the distance it traveled across the floor from the bottom of the ramp against a line of end-to-end sixty inch tape measures fastened to the floor. All results were entered on a chart on the whiteboard and a discussion then followed about the features of the cars that seemed to be associated with the differences between them in the distance traveled. The remainder of the unit was then spent in testing the effectiveness of modifications made to the cars in the light of the hypothesized critical features. Lessons occurred once a week and typically lasted between one and one and a half hours.

Within this basic format, some important changes were made over successive years to enable more systematic exploration of the effect of modifications to the cars. In the second year, following the launch of their home-made cars, each group was supplied with an identical set of prefabricated components, assembled from construction kits bought from a local toy store. Each group’s kit included a Lego baseboard to form the chassis and wheels of different sizes. Nevertheless, there were still problems in controlling all but the particular design variable under consideration. So, for the third year, the engineering department at the local university was commissioned to produce nine identical vehicles that could be systematically modified in various ways. However, even these vehicles did not perform identically until, for one of the later experiments, their wheel bearings were lubricated with oil. These same vehicles were used in the fourth year and, because the effects of the prior lubrication remained, it was finally possible to carry out systematic tests of design variables from the beginning of the unit.

A further change was made in the overall design of the project with the addition, in the second and subsequent years, of a substantial amount of reading and writing. Following the example of Palincsar and colleagues (Palincsar et al. 2001), who had researched the value of guided inquiry learning through the combination of first-hand (practical) investigation and second-hand investigation involving the reading of texts that reported the research of others, we introduced and read from books appropriately written for students at this grade level. In addition, each student was given a research journal and time was devoted each week to students writing a report of what their group had done and discovered. Each week too students were invited to read their entries to the whole class in order to enrich discussion.

While these changes to the overall design of the unit significantly improved the opportunities for systematic inquiry in successive years, the emphasis on students explaining the results of modifications to their cars was present from the beginning. Each week, a substantial period of time was devoted to whole-class discussion of the possible causes of changes in the cars’ performance, which included the forces external to the vehicles and the characteristics of the vehicles themselves.

Data collection and analysis

Data collection was assured by placing a camera with a wide-angle lens in one corner of the classroom so that it captured as much as possible of what was happening during the discussion periods. A wireless microphone was suspended from the ceiling in the middle of the classroom and this adequately captured the contributions of all participants. Photocopies were also made of selected journal entries.

The digital video recording made each week was transferred to a computer and an edited version was prepared that contained only those portions of the lesson that were devoted to
whole-class discussion. These episodes were then coded with the aid of Studiocode, which allows online coding of video data. Selected episodes were subsequently transcribed using the same program.

Coding was carried out at two levels. At the first level, the recording was sequentially segmented into episodes according to the activity involved. An activity was defined as a continuous period of time in which the same participants focused on the same goal; examples include report, experiment, speculate. At the second level, each student’s move was coded for function (e.g. describing, predicting, concluding) and for the function of the move that (a) preceded and (b) followed. Using the same software, transcripts were prepared of those activities that were selected for more detailed analysis.

Studiocode allows the duration of coded events to be calculated; it also prepares a frequency matrix that can be exported to Microsoft Excel. Using both these facilities, the total duration of each activity type was calculated for each lesson as was the frequency of move types within different activities. Because the total duration of whole-class discussion varied across lessons, all results were converted to proportions and, as a second step, averages across lessons were computed for each year.

Findings
The proportion of each lesson devoted to whole-class discussion was very similar in each year; approximately one-third, on average, was spent in discussion (range 33.5–35.2 percent), with the remainder of the time being devoted to building and modifying cars and carrying out trials on the ramp. Discussion time typically involved between three and five types of activities in each lesson, in addition to teacher exposition and management, and these varied from week to week. In the second and subsequent years, most lessons included the activity of students reading from their journals; they sometimes also included time for journal writing. For the purpose of analysis, activities were grouped into five superordinate categories: report included report and describe; speculate included speculate and brainstorm; explain included explain, conclude, and review; literacy/math included read, write, and calculate; and experiment included experiment and predict.

Analysis of the time spent on the various activities showed that much more time was devoted to report in the first as compared with the following years for reasons that will be explained below, while the time spent on literacy/math increased once these activities became a regular part of each lesson in year two. Also notable is the time spent on experiment in the second and third years. Experiments ranged from simultaneously dropping objects of different weights and shapes from the same height to demonstrate the effect of gravity, to rolling balls made of different material and with different circumferences and masses down the ramp to test predictions about the distance they would travel. With a few exceptions, these experiments were performed by Buzz. Interestingly, the proportion of time spent on the activity of explain was 20 percent or more, except in year three, when the lower proportions of both explain and speculate were matched by a greater proportion of time spent on report and reading and writing in journals. These findings will be discussed in greater detail below.

In addition to the varying proportions of time devoted to types of discussion activities across the four years, we also found differences in the frequencies of different types of moves students made during particular discussion activities within a given year. This analysis proved very informative because it revealed that student explanations did not always occur during activities where their articulation was the obvious or stated goal; it
The results of this analysis, shown in Figure 2, illustrate the distribution, in each year, of student moves clustered into five superordinate types along the x-axis in relation to each of the five activity categories which are represented by the five different patterns labeled at the top of the chart.260

In the first year of the project, when controlling variables were not feasible, report (stating actions or events) was a very common activity during whole-class discussions; it was also the activity that included the overall greatest number of student contributions (Figure 2).5 While students did offer a few conjectures and explanations during report activities in 2003, they mainly described modifications they had made to their cars or aspects of their team car’s performance during trial runs. Interestingly, in addition to providing descriptive information, the students in 2003 were apt to offer numerous conjectures about why they thought their car was performing the way it did, and they did this regardless of the activity in progress.

Whereas the first year was largely dominated by student conjectures or descriptions of their observations during report activities, the following year provided more opportunities for students to make explanatory claims that could be tied to the outcomes of their investigations. In comparison with the first year, as shown in Figure 2, the number of student contributions during explain activities in 2004 increased and, looking across 2004 activity types, there were more student contributions that could be categorized as explaining moves. Still in 2004, speculate (open-ended exploratory discussions where students shared opinions and ideas) remained the most common discussion activity, accounting for the largest total proportion of student contributions in any category that year.

The results show that in 2005, the activities report, experiment/predict (choosing among alternative predictions suggested by the teacher), and literacy/calculation (reading from or
writing information in journals or reporting calculations to the class), which did not require much interpretive work or original theorizing, were the activities during which students made the greatest proportion of contributions. Neither speculate nor explain were common activities in 2005, nor did these activities generate very many student contributions (less than two percent of student contributions in any category). By contrast, in all three years other than 2005, the activity speculate was consistently an activity that generated relatively high proportions of all kinds of student moves, and particularly the kinds of exploratory utterances involved in generating and negotiating scientific explanations.

2006 stands out as the only year in which the proportion of student explaining moves exceeded the proportion of other moves during explain discussion episodes, where the organizing pedagogical goal was to connect experimental evidence to claims. Furthermore, the proportion of student explaining moves (11.1 percent) generated during explain episodes in 2006 was just slightly more than the proportion of those generated during speculate episodes that same year (9.5 percent). This is notable in itself; that is, speculate activities in 2006 were successful not only in motivating students to generate the greatest number of hypothetical conjectures/opinions (14.1 percent) in any year but also included a comparatively high proportion of more grounded explanatory comments.

Even as student moves are organized by activity goals, they are also situated in the ongoing stream of discourse. Most whole-class interaction proceeds through sequences of three-part exchanges, i.e. initiation-response-follow-up. In this format the teacher’s initiating question typically calls for information that the students are expected to know already and the follow-up move evaluates the correctness of the response. Since this pattern tightly constrains the form that student contributions can take, two changes are necessary for any form of dialog to develop. The first is the choice of the opening move, which needs to be in a form that invites a range of possible responses, e.g. ‘what do you think might be the reason for . . . ?’ One strategy Buzz used in this respect was to ask an open-ended question and encourage several volunteers to respond, accepting each, irrespective of his opinion of it. Even more important is how the student’s response is treated; this means that the follow-up move, if one is made, should communicate interest in what has been offered, building on it, asking for more information, or in some other way showing ‘uptake’ (Nystrand 1997).

As seen in Table 1, Buzz used a variety of follow-up moves that gave recognition to students’ ideas: He checked to make sure he had understood, reformulated a contribution more succinctly, asked for the reasoning behind a suggestion, summarized the alternatives, and so on. As the project developed, he added other types of moves to his repertoire, including that of giving no follow-up at all.

Using a taxonomy of types of follow-up, we coded the sequel to every student contribution across all four years. The results are shown in the top part of Table 1. Several findings are worthy of notice. First, evaluating moves were very infrequent, constituting less than one percent every year. Second, the frequency of checking or reformulating student contributions was greatest in 2005, the year in which a large increase in the proportion of ELLs occurred. And third, simple acknowledgement or the complete absence of follow-up (null) tended to increase over the years, making the interaction more conversational in nature. Also worthy of note, as seen in the lower half of the table, is the relatively high proportion of unsolicited ‘offers’ of information by students in 2004 and 2006, both of which years were judged overall to be the more successful years, and in particular, in 2006, the much higher frequency of students responding to other students, thereby making a teacher follow-up unnecessary. Thus, although Buzz controlled the overall organization of the discourse through the moves he made in initiating new topics, the general tone was
Table 1. Distribution of follow-up moves per year.

<table>
<thead>
<tr>
<th></th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Teacher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acknowledge</td>
<td>10.9</td>
<td>6.6</td>
<td>10.6</td>
<td>13.2</td>
</tr>
<tr>
<td>Evaluate</td>
<td>0</td>
<td>0.6</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Counter</td>
<td>3</td>
<td>4</td>
<td>1.7</td>
<td>5.2</td>
</tr>
<tr>
<td>Check</td>
<td>19.1</td>
<td>5.7</td>
<td>14.9</td>
<td>11</td>
</tr>
<tr>
<td>Repeat/reformulate</td>
<td>9.7</td>
<td>8</td>
<td>18</td>
<td>11.7</td>
</tr>
<tr>
<td>Comment</td>
<td>8.6</td>
<td>6.8</td>
<td>11</td>
<td>6.7</td>
</tr>
<tr>
<td>Praise</td>
<td>2.2</td>
<td>2.8</td>
<td>1.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Uptake</td>
<td>5.2</td>
<td>7.4</td>
<td>3.3</td>
<td>8.2</td>
</tr>
<tr>
<td>Follow-up question</td>
<td>12</td>
<td>7.1</td>
<td>8.7</td>
<td>8</td>
</tr>
<tr>
<td>Null</td>
<td>5.2</td>
<td>5.7</td>
<td>3.9</td>
<td>7.2</td>
</tr>
<tr>
<td><strong>Student initiations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student replies to student</td>
<td>23.2</td>
<td>10</td>
<td>3.1</td>
<td>46.4</td>
</tr>
<tr>
<td>Student offers</td>
<td>4.9</td>
<td>26.8</td>
<td>4.6</td>
<td>23.2</td>
</tr>
</tbody>
</table>

remarkably dialogic in nature, given the number of participants involved (Wells and Mejía Arauz 2006).

Our arguments below focus on how, in each year, the construction materials, other apparatus, means of representation, and Buzz's changing stance with respect to the practices and concepts organizing the unit mediated students' contributions to the discussion and thus their emerging understanding. However, we also recognize that some of the variations from year to year were contributed by the different individual students involved; because each student brings to the classroom her or his own individual history, and all the different commitments, assumptions, positional identities, values, motivations, skills, etc., derived from their past experiences, all of these individual characteristics, inevitably shaped the classroom interaction that we observed.

**Students’ developing understanding of factors affecting car performance**

Every year students were strongly motivated to understand what factors affected the performance of their cars. In the first two years, discussion was mostly focused on design factors as they attempted to improve those features of their vehicles that were negatively affecting the performance. In the last two years, on the other hand, there was more concern with the external forces that might affect performance, most notably gravity and friction. In all four years, however, there was much discussion of the influence of the weight of their vehicles. So, due to limitation of space, we shall confine our discussion here to the ways in which class members in successive years developed an enhanced understanding of the influence of weight.

In 2003, because of the diversity of materials, the size and the weight of the cars being built were uncorrelated. Yet, in an early lesson the effect of weight was raised by Martin when he compared the performance of his group’s car with that of another group’s car.

Martin: I think our car ran really well but um I think we should make some adjustments . . . and I was thinking maybe it was because . . . theirs was a little bit heavier maybe or . . . I THINK – I don’t have xxx I just think that maybe our wheels weren’t aligned up very well ’cos we -

T: You’re talking about the car that went into the seat?
T. Ball and G. Wells

Martin: Yeah. so I think – that’s why I think we should go – maybe make it a little bit heavier.

Weight also seemed important to Sophie. Commenting on her group’s car a moment later, she proposed the following:

Sophie: I think it should either be light or heavy. I think what we’re dealing with – I think if it’s light it sort of has something – it doesn’t have to have much to push, but if it’s heavy it has more to push down. So I think both of them [light and heavy car designs] are pretty good. but I think it [weight] makes a difference.

These brief mentions illustrate both the design (vs. scientific) orientation of the car builders and also the way in which reporting on their cars’ performance often led to conjectures about potentially critical design factors. Weight continued to be considered a likely variable, but much of the available evidence was inconclusive because the variability of the materials and assembly of cars resulted in disparate and, at times, seemingly contradictory results.

At the end of the unit, when the teacher asked the students to draw conclusions about their experimentations, there was still ambivalence about the effect of weight on cars. As Martin pointed out, the car that traveled the furthest was heavier than most, but the second longest distance was achieved by the lightest car. Recalling the early hypothesis, supported by several students, that adding weight would make cars go farther, Buzz called for a show of hands. This revealed that three out of seven groups had confirmed the hypothesis, but another three had failed to do so. The only possible conclusion, therefore, was that no conclusion could be reached about the effect of weight.

By the second iteration, in 2004, new materials were acquired with the intention of making it more possible to carry out systematic tests of selected variables. In one of the early lessons, when students were reading from their journals about their discoveries so far, Raoul interjected, ‘I think that the lighter it is, the farther it goes’. Acknowledging the significance of this conjecture, the teacher replied, ‘We’re going to get into some of those things today’.

The following week, Buzz started the lesson by putting three groups’ cars on a table and asking the students whether, just by looking at them, they could tell what accounted for their performance.

T: How do you know for SURE – without a doubt – what made it go the farthest?
Christie: I think xxx
T: You THINK because it’s lighter. it might be . . .
[Several other students offer conjectures]
T: You all said you ‘think’ because you’re really – you can’t be sure because of the variables. there are a lot of variables. could it be the wheels? could it be the chassis? – they’re all pretty different. could it be the weight? . . . could it be the axles?

Following this discussion, it was decided to carry out a controlled experiment. We think it is significant that this decision was reached through consensus rather than imposed by the teacher; the class based their decision on the shared understanding they had gained of the problem posed by the fact that the cars were all so different. Of all possible variables, weight was the easiest to test because additional weight could easily be added to any car,
irrespective of its design. Keeping all other design features constant, every group ran trials with and without added weight. Checking the results the following week, it was found that there was no car that had traveled farther when weight was added. As one student observed, ‘It doesn’t help. It’s worse’. This finding was unexpected and immediately gave rise to a host of explanatory conjectures.

Bodey: Well I think it’s the weight . . . like when it hits the flat on the – on the carpet . . . it like sorta like is a weight its like it like umm . . . the um the ah chassis – we don’t have one on ours – but all that connects (indicating plastic connector rods) – the pennies is like pushing it DOWN so like the wheels kinda STOP and slow down.

T: So the weight coming down the ramp – pushing your car down – as soon as you hit the flat . it kind of pushes it DOWN . it’s not pushing it down the ramp any more, it’s weighting it down on the carpet that slows the wheels down?

Toward the end of the lesson when this conjecture was repeated, the teacher asked why adding weight should have this effect; when no explanation was given, he demonstrated sliding his hand across the carpet with more or less downward pressure and commented on how much more difficult it felt in the latter condition. Immediately someone suggested friction to be the problem and the teacher agreed that friction was something that needed further investigation.

In the following weeks, the issue of friction was explored in two ways. First, the wheel bearings on each car were lubricated with oil. This resulted in all cars traveling greater distances, with and without weight. Second, it was decided to try running the cars on different surfaces based on the hypothesis that some surfaces might create more friction than others. In practice, the most convenient alternative was the asphalted school yard, which was expected to cause less friction than the classroom carpet. Accordingly, the ramp was set up at the same angle outside and each group ran several trials. Unfortunately, the asphalt yard was not completely level and individual results depended on whether or not the car encountered a slight incline. Nevertheless, the general conclusion based on all results was that cars traveled further on the asphalt than on the carpet and this was because there was less friction on the asphalt. However, because of the variability in paths traveled, no conclusion could be drawn regarding the effect of added weight on the asphalt.

Nevertheless, one student believed he had understood the relationship between weight and friction on the asphalt. In a discussion that followed the time spent outside, Jerry offered the following explanation.

Jerry: Well, er – . this is my theory of why weight helps when you’re on the asphalt-

T: What do you mean, theory?

Jerry: My theory like . well it’s not really xxx but it’s like weight – why weight causes x the asphalt – cos the asphalt has little bumps which xx – the little bumps make the cars go UP a little bit and that slows them down . but with weight um it keeps the car um going to – going THROUGH the <bumps> like the bumps aren’t even there

Despite the increased understanding evidenced by the weekly discussions, it is clear that, in the second year, weight was still being considered mainly from the design perspective: Would adding weight cause the cars to travel further? Nevertheless, the students did gain a practical understanding of the need to control all but one variable when conducting an experiment; they also learned the importance of averaging the results of several trials to
T. Ball and G. Wells

arrive at the most accurate estimate of the effect of a chosen variable. Although the class as a whole spent more time speculating about the observations they had made rather than drawing firm conclusions, slightly more explaining did emerge in 2004 and individuals were able to come up with feasible explanations like Jerry’s theory above. In 2003, by contrast, students rarely moved beyond straightforward descriptions and tentative conjectures. On the other hand, while the term ‘momentum’ was used several times by students and adults during the second year, at no point was a clear connection was made between the more abstract concepts of weight, gravity, momentum, and friction. The following interchange was probably the most coherent statement of the relationship.

Jerry: I think when the ramp is steeper it [the car] has more force going down and with more force it has more speed. that way it can go farther on the carpet

T: What is that force?

Jerry: Gravity

The third iteration in 2005 started, as in previous years, with work on cars students had built at home and each week time was allocated for discussing the changes they had made and why. However, this year saw the more systematic use of weekly journal entries. As can be seen from one student’s entry in the first week, weight was quickly recognized to be an important variable.

I think heavier cars will go faster because gravity pulls down weight so the more weight the more gravity will pull it down making it go faster

Several students acted on the same assumption and added weight to their cars. In most cases – barring wobbly wheels – this assumption proved to be borne out by the distance traveled. However, other possible features, such as size of wheels, texture of wheels, and how easily the wheels turned on the axles, were also suggested. Clearly, at this point, the students were thinking as engineers focused on results and there was no further mention of the forces involved. So, in the following week, to explore the possibility that multiple factors might be responsible for the variable distances different cars had traveled, balls of different diameter, weight, and texture were run down the ramp; and in response to the results of this ‘experiment’, the class agreed that clearly no single factor was solely responsible.

At this point the cars commissioned from the university were introduced with the intention that these should be used to carry out systematic ‘scientific’ tests of the proposed variables. To prepare for these tests the teacher elicited different aspects (length of chassis, size of wheel, weight, etc.) that students thought might make a difference and listed them on a chart posted on the whiteboard at the front of the classroom. The concept of a ‘fair test’ was emphasized, together with the need to take the average of three trials as the best estimate for each value. Following the students’ interest, weight was the first variable chosen. Cars were run in their current state, then with the addition of a small weight, and finally with successively larger weights. Students were asked to write down their predictions in their journals, keeping in mind the other variables listed in the chart. The result, discussed the following week with reference to a graph of average distances covered, was that initially adding weight made the cars travel further, but beyond a certain point additional weight had no effect. Pondering over this unexpected result led to a consideration of the negative effect of friction and so to a practical comparison of running the cars in the classroom with its carpeted floor and out on the asphalt in the school yard. The results of this comparison were quite clear: Adding weight in the classroom caused the cars to cover less distance, but on the asphalt, it caused them to cover a greater distance. Finally, the axles were oiled and
it was found that the cars traveled farther under all conditions, in fact, so much so that it
was necessary to reduce the angle of the ramp to avoid the cars hitting the wall at the end
of the classroom.

The last lesson of the unit started with the students writing about what they had learned
in the preceding weeks. Not surprisingly, given the recent experiences with adding oil and
running the cars on different surfaces, friction received the greatest amount of attention.
Nevertheless, there was also a strong and continuing interest in the effect of weight and its
interaction with gravity and friction. While reading what he had written, John recalled the
unexpected result of increasing the weight of the cars, and offered the idea of there being a
‘peak weight’. This led to the speculation that beyond a certain point the addition of more
weight increased friction on the ramp with the result that the cars did not travel down fast
and therefore gained less momentum.

In 2005, the use of the commissioned cars (once the axles had been oiled), the regular
interpolation of pointed ‘demonstration/experiments’, more systematic directions guiding
student journal entries, and Buzz’s own developing clarity around the physics involved,
combined to make it more possible to reach decisive conclusions about the effect of
adding weight under different conditions. However, as pointed out in the discussion of
our quantitative results shown above, this was also the year when students made overall
fewer individual contributions, and the contributions they did make were more likely to
be descriptive reports rather than theoretical or explanatory answers to why questions. We
believe that these findings can be explained, at least in part, by the dramatic changes in
the demographics of the class due to changes in school zoning laws that took effect that
year (described earlier), and second, by changes in Buzz’s own stance toward the project
from the kind of exploration motivated by genuine curiosity and perplexity in the beginning
years to lesson designs based on increasingly firm convictions regarding content goals.

In 2006 ‘weight’ was regarded as a crucial factor from the outset. By the time this fourth
iteration of the unit came around, the incoming fourth-grade students had caught wind of
Buzz’s annual car ‘contest’, either in the schoolyard or because they or their friends had
older siblings that had participated in a previous year. Invested in building a car that would
outdistance others on its first run down the ramp, on the first day of the unit several students
(sometimes with their parents) brought in very hefty cars, incorporating components like
lawn-mower wheels or a block of concrete.

Partly to shift attention away from the unit’s growing reputation as a contest (rather
than an inquiry) and partly in response to students’ focus, in previous years, mainly on
factors like wheel size, type of axle, and weight rather than interactions between these
and underlying ‘invisible’ processes like gravity, friction, and inertia, Buzz posted a new
caption on the ramp: ‘May the forces be with you’; this was meant to serve as a subtle but
continuing reminder to the students that there was more to this endeavor than the ‘finish
line’. Strategically, the second day of the unit was spent entirely in discussion rather than
on construction or trial runs. Buzz used two strategies to give impetus to the unit and to
scaffold a focus. For one, he began the discussion that day by reading aloud a version of
Newton’s first law of motion.

T: ‘An object will remain at rest’ – at rest, that means it will just be sitting there – ‘or will
continue moving at the same speed and in the same direction unless a force acts on it’.

Next, drawing on Palincsar and colleagues’ (2001) use of a fictional lab notebook, he
introduced a fictional personality: Cynthia Clark, an undergraduate scientist, and proceeded
to read journal entries from, what he told the class, was Cynthia’s lab notebook, which he
had on loan. He explained that he wanted to share Cynthia’s comments because he had learned that she just happened to be investigating the same kind of questions that they were, such as ‘what makes things go down the ramp?’. Buzz had copied ‘excerpts’ of Cynthia’s notes and included them in folders for each student. Asking the students to follow along, he related how in the first place Cynthia got interested in these questions while reminding the students that, like many scientists, she had made a habit of taking out a notebook to write down her thoughts.

T: It’s really interesting how she got started in it, and that is, over here by Depot Park, the skate-park . . . she was watching all the skateboarders do all these things, and she didn’t know why, so she went home and built a little ramp like ours and tried to figure it out.

As he read aloud, Buzz made a point of connecting what Cynthia was doing to the investigations they had already been pursuing in their classroom.

T: (reading aloud) ‘Sunday I was out riding my mountain bike . . . when I came to the skatepark near Depot Park . . . I was watching the skaters do some amazing tricks and maneuvers. While I was watching I noticed that some seemed to go faster than others, and I started to wonder about some things . . . ’ This reminds me of your cars, some of them went down faster – and it reminds me of the balls [referring to an in-class experiment], some went faster than the others . . .

At this point, in response to Cynthia’s musings, a number of students had their hands in the air, so anxious to add their own comments that many could hardly refrain from interjecting. Staving them off for a longer moment, Buzz finished reading the first of the series of Cynthia’s questions before opening up the dialog:

T: Hold on – let me just see what her questions were first: (reading aloud) ‘Number one: why do some skaters seem to go faster [down the ramp] than others?’ I don’t know, do you guys want to take that one on?

The teacher’s strategy of connecting Cynthia’s questions both with the students’ collective experience in the classroom and with a locale (the skatepark) that was, to varying degrees, part of the students’ everyday lives outside the school, seemed to work. In the ensuing brainstorming episode, students could hardly wait to have their say, often talking to each other, flooding the airwaves with a multitude of suggestions about the different materials used to make skateboards and skateboard wheels, the diameter of the wheels, different types of bearings and pegs, the shape of the skateboard, etc. While encouraging and accepting all the ideas students had, Buzz made repeated and concerted efforts to highlight conjectures that alluded to an underlying force.

T: Were there any other reasons that you guys thought that some skaters go down faster than others?

Benji: What, when they go down the ramps?

T: Yeah why do some go faster than others?

Benji: Oh, oh, it was, I know why . . .

Andy: People’s weight!

Trevor: The person that’s riding it-
T: Hang on, so the rider, him or herself, whatever they weigh, might make some difference to whether they go faster or slower?

Linda: Or the materials of the board

T: Right the material of the board, we got that-

Benji: And – and how many times, how many times they kick their foot on the ground to ride.

T: Oh so the push-off they get when they start might be stronger –

Andrew: Yeah, the force.

T: What did you say? oh force . ‘force’ well, there’s a cool word . the force at the start might be different.

Tyler: How high the ramp is at the start that they are going off-

T: Oh if they are on different size ramps, okay, I hadn’t thought about that . I was thinking they were all on the same one, but that would make a difference

This animated conversation generated all kinds of issues for further exploration by the class, and weight was clearly viewed as only one of many important contributing factors. But before the class moved on to investigate different scenarios proposed by the students, the teacher introduced the set of cars commissioned from the university. He emphasized that although all cars appeared to be manufactured in the same way, still they couldn’t know for sure whether all seven cars would perform identically.

T: How would we know that all the cars are exactly the same today? There might be a difference so that they don’t all go the same... we’re trying to find out how to test these cars to figure out if they are all pretty much the same right now today

True to the spirit of collaborative inquiry, the decision to establish a baseline distance measurement for each car came, after some speculation, as a suggestion from a student rather than as a mandate from the teacher. This was typical of the way in which the class arrived at many of the decisions they had made about how to proceed in their investigations. Buzz’s democratic approach not only made students feel that their suggestions were valued but, by creating conditions under which students were compelled to articulate their arguments, he created scenarios that led the students to a more reflective understanding of the scientific practices they were involved in executing.

Overall, the 2006 experimental results were much more conclusive than those of the first two years, since the commissioned cars, now with oil on the axles, allowed the class to control a chosen variable more carefully. This meant that it was more feasible for the teacher to initiate activities where the goal was to relate claims to evidence as compared with previous years, when disparate results limited the kinds of claims students could make about the outcomes of different trial runs. By the end of the unit, the class had concluded decisively that the cars traveled farthest outside on the asphalt with added weight. By contrast, with added weight inside on the carpet the cars traveled the least distance. These results set the stage for a productive discussion exploring the intersections between gravity and friction. Meanwhile, positioning Cynthia the scientist as a ‘pen pal’ who was interested in the same issues that were being explored in class worked to create a personalized audience, or sounding board, for the students’ emerging ideas. As is apparent from the quantitative results discussed earlier, in 2006 the activity explain thus became a more feasible and regular activity and more student explanations were generated during explain and speculate activities than in any other year (see Figure 2).
Discussion: issues in the evolution of the project over time

As explained in the introduction, this project was undertaken as a collaborative inquiry on two levels. On the first level, the intention was, with successive groups of students, to study their developing understanding of the interplay between design features and the physical forces involved in accounting for the distances covered by cars when released down a ramp. On the second level, our purpose was to discover how far and in what ways the first investigation would promote the development of dialogic inquiry and the attempt to arrive at warranted explanations of observed events. In the previous sections, we have focused on the students' engagement in inquiry and in their developing understanding of the physics of motion. This involved their learning about the practicalities of carrying out fair tests as a prerequisite for the conduct of scientific investigation and learning to make intelligible conjectures and predictions in relation to changes to be made to cars and the 'test track', and to construct explanations of the results obtained from controlled trials. As has been shown, each year was different in the amount and kind of progress made, although in all four years this was substantial for the majority of students.

What is needed to complete the picture is an account of what the adults learned and the changes that we attempted to make as a result. However, since most of the changes and the reasons for them have already been described in the preceding sections, and because of limited space, we shall restrict our discussion to a small number of problematic issues.

Science or engineering design

Given the adults' intention that the project should provide an opportunity to introduce students to the (invisible) physical forces affecting their cars' performance, the impossibility of carrying out tests of the effect of differences in weight and in the friction of the track surface was somewhat frustrating. It was largely to overcome this problem that the more standardized materials were provided in the second year and that identical vehicles were commissioned in the third year. However, despite absence of clear results in the initial year, the students were not deterred from making conjectures about potentially critical variables and, in the case of some groups, from modifying their cars in an attempt to test them. However, their orientation was very much that of engineers who were trying to 'win' the competition to have their car go farthest.

By 2005, the problem of incommensurability of group-designed cars was overcome by the introduction of the commissioned cars. In addition, considerable time was spent by the adults in explaining the need for fair tests and encouraging the students to make the connection between weight, gravity, momentum, and friction. However, the opportunity to speculate about these relationships and to explain trial results in their terms were taken up to a lesser extent than in the previous year (2004), when the differences between student-built cars still made controlled tests difficult to carry out. (To some degree, the unexpected variability in performance among the commissioned cars before they were oiled may have had a dampening effect on that year's students.) By contrast, in the final year (2006), as can be seen in Figure 2, student conjecturing and explaining moves once again occurred with high frequency.

In many respects, this fourth year could be judged to have been the most successful in meeting the initial aims of the project, due in part to different ways in which the unit started. However, in the last lesson, in which the class reviewed the work they had done and what they had come to understand, when asked how the unit could be improved, one student suggested that more time should be spent on using what they had learned to improve their
home-made vehicles, a suggestion that was supported by several others. This could be interpreted as a continuing preference, on the part of students, for being engineers rather than scientists, or at least for being applied scientists.

This issue was discussed at some length during our interview with Buzz. Seen from the adult perspective, we had learned a great deal about the effect of gravity on an inclined plane and about various ways in which friction reduced momentum. As Buzz explained, each year during the unit he would come to school really early and carry out experiments as a way to narrow possibilities, largely to satisfy his own curiosity, and also for the benefit of the students in helping them to resolve some of their confusions. Furthermore, as he became more knowledgeable, he also became somewhat more directive in deciding what activities to include in each week’s lesson. From the students’ perspective, however, although they achieved a better understanding of the relevance of the physical forces, these forces were not open to direct manipulation in the same way as the design features of their cars; and since making their cars travel as far as possible was the students’ ultimate aim, the physical forces were to be utilized or overcome rather than treated as of interest in their own right.

As we discussed the apparent conflict between the design and the scientific stance, we realized that our relative success in getting the students to focus systematically on the ‘science content’ was bought at the price of their enthusiastic discussion of the conjectures that occurred when the results of trials could be attributed to several alternative design variables or to a combination of them. In these situations, the resulting confusion several times led to valuable ‘ah-ha’ moments for some students. In Buzz’s view, they were both important. We had some of those moments earlier in the year when they brought their own cars that they’d built at home. They had all kinds of ideas about what was happening. Also those moments where I’ve just put this out here and there’s some confusion . . . [but we also need] some really prescribed ‘here’s what I think is going to happen ‘cos we’ve done all of this . . . and so I think a combination . . .

Increasing participation

As a class teacher, Buzz was very conscious about the tendency of participation to become unequal during class discussions. ‘One of my goals was always to find out ways to get more kids engaged and involved in what we were doing’. This problem was made more acute by the much higher proportion of ELLs in the class in 2005 and the subsequent year. Several changes were made in order to try to ensure that no student would ‘go underground’, as he put it. The greater degree of teacher direction was one such change, as was the practice of having students read to the class what they had written in their research notebooks about the activities in the previous lesson. The introduction of Cynthia and her research on skateboarding in 2006 was another attempt to harness a known interest in students and the opportunity to correspond with her certainly seemed to prompt many students to think more deeply about what they were doing.

Throughout all four years, we tried to let the students take the lead in deciding on the line of investigation to follow; similarly, we waited for them to introduce the more technical terms, such as ‘gravity’ and ‘momentum’, though we were quick to show uptake when they occurred. On viewing a clip of one such occasion, Buzz commented,

I think we kind of made the agreement among ourselves from the very beginning that we weren’t going to make any suggestions like that unless they brought things up. In fact, I think we all got better at questioning – directing them and guiding them to get them questioning about what things were happening.
Nevertheless, despite our best attempts, there continued to be a small proportion of students who remained reluctant to participate in discussion or were easily distracted by other topics of conversation of a more personal nature. This was exacerbated by the fact that, when working in their groups, only one or two students could be directly involved in making modifications to the cars or play a central role in carrying out the trials. Buzz's tighter planning of lessons in the later years with the inclusion of demonstration experiments and the practice of writing and reading from their individual journals was, in part, a managerial strategy designed to keep everyone engaged. There is no doubt in our mind that the alternation of action and more reflective discussion was, in general, successful in sustaining interest. But perhaps we were too ambitious in believing that students of this age were ready to be as interested in what to them were abstract ideas as much as in the excitement of 'winning the race'.

Developing dialog

In our view, engaging students in self-directed practical inquiry is essential to enable them to generate ‘wonderful ideas’ (Duckworth 1987), but collaborative dialog is equally essential for the clarification and development of understanding. However, for such dialog to occur, a change of orientation is often required on the part of the teacher, for traditional teaching allows little opportunity for students’ ideas to be heard – for their opinions to be invited and taken seriously. From the beginning, Buzz was aware that ‘I needed to find ways to stimulate and get them interested so that they’re going to get this conversation going instead of the lecture-type thing in which four or five kids are interested and the rest of them are sitting there kind of bored.’ And, as he made clear, ‘It’s hard work for [the students] and for me [to] make it be productive and meaningful for them’.

In the light of previous research we formulated some tentative principles that we thought could enable classroom communities to restart the dialog. Over the course of the four years, as we modified the organization of this unit, these principles were largely confirmed. In our experience, then, in order to engage in productive and inclusive dialog the following conditions must prevail:

- The class engages in a shared inquiry.
- The topic under discussion is, or becomes, of interest to the participants, where interest is most likely to be generated when the discussion bears on a future action to be carried out or on an ‘object’ that participants are constructing or trying to improve (e.g. a model, performance, text, or explanation).
- Individuals have opportunities to contribute opinions, suggestions, observations, or experiences that they want to share and believe to be related to the activity in progress.
- Others are willing to listen attentively and critically.
- There are opportunities for all participants to discuss whether, and in what ways, different contributions are relevant.
- Experts share control and the right to evaluate with novices.

Conclusion

In setting out the theoretical basis for our investigation, we argued that learning takes place through a continuing spiral of knowing as, over time, current understanding derived from past experience is tested, critiqued, and enhanced as students engage with new information through collaborative knowledge building to produce and improve explanations with respect
to the object they are working on. In this paper, we have tried to communicate both the process and the product of engagement in this spiral by successive years of students, the teacher, and researchers. While we believe our findings may be informative for other educators, our hope is that this account will persuade them of the value of conducting their own investigations of learning over time.

Acknowledgements
We wish to express our gratitude to Buzz Gray and his students at Shoreline Public School for their contributions to this paper.

Notes
1. Keil goes on to emphasize explaining as a situated practice that is part of a larger social dynamic: ‘The coordination of claims and evidence is not simply a cognitive skill that signifies that data or the concepts underlying claims have been understood but is part of a broader social practice used to persuade other people. Students’ explanations therefore reflect their (possibly tacit) ideas about what makes an argument persuasive to a particular audience’ (2006, 227).
3. Here and in what follows, ‘Teacher’ also includes the researchers, although this role was taken up by Buzz approximately two-thirds of the time.
4. The full set of coding categories can be found in Appendix 1.
5. As might be expected, different activities elicited different types of moves; consequently, not all move types occurred in every activity.
6. We also believe that the processes through which students came to understand issues related to weight are fairly representative of the ways in which other concepts were developed and applied.
7. With this discussion in mind, we prepared a selection of video clips from each year to provide a starting point for an informal interview with the teacher, which was audiorecorded.
8. In fact, this suggestion was taken up in the following year.

References
T. Ball and G. Wells


Appendix. Coding categories used for analysis

*Activity*: Brainstorm, experiment, predict, describe, suggest, intend, speculate, explain, conclude, review, read, write, calculate, teacher exposition, teacher management.

*Student move*: Informing, observing, describing, suggesting, intending, predicting, conjecturing, explaining, concluding, questioning, offer, respond to student.

*Follow-up*: Acknowledge, evaluate, check, repeat, reformulate, comment, counter, uptake, follow-up question, praise, null (= no follow-up).