



HAL
open science

Peer argumentation in the school science laboratory - exploring effects of task features

Per Morten Kind, Vanessa Kind, Avi Hofstein, Janine Wilson

► To cite this version:

Per Morten Kind, Vanessa Kind, Avi Hofstein, Janine Wilson. Peer argumentation in the school science laboratory - exploring effects of task features. *International Journal of Science Education*, 2011, pp.1. 10.1080/09500693.2010.550952 . hal-00692132

HAL Id: hal-00692132

<https://hal.science/hal-00692132>

Submitted on 28 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Peer argumentation in the school science laboratory –
exploring effects of task features**

Journal:	<i>International Journal of Science Education</i>
Manuscript ID:	TSED-2010-0242.R2
Manuscript Type:	Research Paper
Keywords :	argumentation, laboratory work, science education
Keywords (user):	

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Peer argumentation in the school science laboratory –
exploring effects of task features**

Prepared for International Journal of Science Education

For Peer Review Only

Peer argumentation in the school science laboratory – exploring effects of task features

Peer argumentation in the school science laboratory – exploring effects of task features

Abstract

Argumentation is a significant component of scientific inquiry: introducing these skills into laboratory work can be regarded as a goal for developing practical work in school science. This study explored the impact on the quality of argumentation among 12 – 13 year old students of laboratory-based tasks of three different designs. The tasks involved students collecting and making sense of complex data; collecting data to address conflicting hypotheses; and, in a paper-based activity, discussing pre-collected data about an experiment. Significant differences in the quality of argumentation prompted by the tasks were apparent. The most argumentation units per unit time were generated by the paper-based task. Where students had to carry out an experiment, argumentation was often short, as reliance on their data was paramount. Measurements were given credence by frequency and regularity of collection, while possibilities for error were ignored. These data point to changes to existing practices being required in order to achieve authentic scientific inquiry in school laboratory work.

*Peer argumentation in the school science laboratory – exploring effects of task features***Introduction**

A rapidly growing interest in *argumentation* among science educators has been fuelled by socio-constructivist theoretical frameworks in science philosophy and learning psychology (Jiménez-Aleixandre & Erduran, 2008). In science philosophy, science theories are no longer seen as products of pure empirical or logical processes but as ideas shaped through critique, debate and revision within the science community (Kuhn, 1962). In learning psychology, learning is seen as originating from socially mediated activities (Vygotsky, 1978). Knowledge and cognitive processes exist in a social milieu and are internalised by the individual through language and active participation. From these key ideas it has been claimed that science education needs to change from a current “positivist” practice, emphasising a misleading picture of science and factual recall of knowledge, towards a practice that helps socialisation of young people into the norms and practices of authentic science (Driver, Newton, & Osborne, 2000). The current paper relates these changes to teaching and learning in the science laboratory. To date, socio-scientific issues have provided a more common focus for argumentation in science education activities, perhaps because preparation for decision-making in political and moral issues students meet in society (Zeidler, Sadler, Simmons, & Howes, 2005) has dominated educational thinking. Laboratory activities may be an attractive alternative drawing the attention towards argumentation in a scientific context and demonstrating something about the architecture of scientific knowledge (Ford, 2008). Osborne, Erduran and Simon (2004) offer a second reason for the dominance of the socio-scientific context, by showing that initiating argument in a scientific context is much harder. To keep argumentation meaningful students need understanding of the evidence (Koslowski, 1996), which can be challenging for some scientific phenomena. In this view, the laboratory also may be favourable because students familiarise themselves with the phenomenon and generate their own data.

There are, however, obvious reasons that argumentation in the laboratory is difficult (Hofstein, Kipnis, & Kind, 2008). A typical school inquiry more closely resembles the empiricist-logical view of science that dominated science philosophy a century ago than today’s socio-constructivist view (Driver et al., 2000). Scientific method is presented in a step-wise manner leading students steadily along a path from research problem to final conclusion. Thus, they are led in ways that avoid the complex nature of authentic science in order to ensure that the “right” data appear and the “right” conclusions are being made. Open-ended investigations with more degrees of freedom occur, but are limited in their use for the pedagogical challenges they present to the teachers (Hofstein & Lunetta, 2004). Restrictions are also made by the need to reach curriculum targets and conform to assessment practices (Donnelley, Buchan, Jenkins, Laws, & Welford, 1996). Breaking with these practices is a major challenge if we want to establish the science laboratory as an efficient place for teaching and learning scientific argumentation.

Against this background the current paper presents an exploratory study that has investigated the effect of specific pedagogical strategies in laboratory tasks. Formats of laboratory tasks have been analysed extensively by Chinn and Malhotra (2002a) and they

Peer argumentation in the school science laboratory – exploring effects of task features

define one group of tasks as “simple experiments”, which has been the main focus in the study. These are tasks in which students typically are given a research problem to investigate the effect of one or two provided variables, and with some openness in the choice of method. The study is set in a UK context, where such tasks are referred to as *investigations*. By suggesting particular pedagogical strategies implemented in the tasks and trialling these in “ordinary” teaching contexts and accepting that classroom life is synergetic (Brown, 1992), we hope to shed some light on why argumentation in the science laboratory is a challenge and what specific actions may be taken to improve current practices. Before introducing the study the paper reviews research analysing discourse practices in the science laboratory and sets out a theoretical rationale.

Argumentation in the school science laboratory

Abrahams & Millar (2008) contrast two levels of efficiency in laboratory teaching. On the level of having students “do” practical activities, the teaching is efficient: if we enter a science laboratory we will mostly find well-functioning teaching with students conducting activities as expected by their teacher. This efficiency, however, is not reflected at the level of students’ learning. Abrahams and Millar observed twenty-five laboratory lessons and found that the teaching was generally ineffective in helping students achieve learning as intended. The reason, they point out, is that most time is spent on manipulating equipment *rather than* discussing ideas. This is supported by Newton, Driver and Osborne (1999), who also observed a series of randomly selected lessons. They found as little as 0.4% of time spent on laboratory tasks was organised as group discussion. Practical lessons were generally found to include *less* argumentation than non-practical lessons. Watson (2004) offers a more encouraging picture, focusing uniquely on open-ended investigations. He observed a laboratory task running over three lessons and found students frequently making subject-related claims and statements. The argumentation, however, was simple because students met claims with counter-claims rather than supporting data. Similar findings are also reported in a study by Katchevich, Mamlok-Naaman and Hofstein (2010) in Israel. They found that although students made simple arguments, open-ended investigations in chemistry made significantly more opportunities for argumentation than closed experiments. Arguments developed when students were building hypothesis, analysing results and drawing conclusions. Most positive results on teaching scientific argumentation in the laboratory, however, are reported from studies more actively using strategies “forcing” students to present and debate claims. In Korea, for example, Kim and Song (2006) stimulated argumentation in an open-ended investigation by having student present reports for “peer review” and by arranging discussions in which students acted as critics in a similar manner as scientists on a conference. By this arrangement they were able to reach a much stronger focus on argumentation than in the previous mentioned studies. Similar findings are presented by Kolstø and Mestad (2005) in Norway. They found that the strategy of peer reviewing laboratory reports worked as a natural way of stimulating scientific argumentation, because students in this way were led to defend and question conclusions in light of gathered data.

Theoretical rationale

The above studies suggest traditional laboratory teaching has established a pattern of behaviour that prevents argumentation, and that changing this requires use of specific

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3 teaching strategies. Strategies proven most efficient have a role-play nature, making
4 students act the role of scientists. However, even if this is a positive contribution, we do
5 not think it sufficient. Not all laboratory activities can be arranged as role-plays and we
6 should be looking for a more general change of the teaching and learning culture.
7

8
9 Classroom cultures supporting appropriation of argumentation skills have been much
10 discussed and analysed with focus on social and epistemological perspectives. The social
11 perspective points towards the need for a climate that fosters dialogic discourse (Zohar &
12 Nemet, 2002). Small-group discussion is a commonly used technique; which seems
13 natural in laboratory teaching, but has had little success (Newton et al., 1999). Reasons
14 for this may be lack of teacher encouragement and specific procedural guidelines
15 (Herrenkol, Palincsar, DeWater, & Kawasaki, 1999). The teaching needs to actively
16 encourage mutual student-student and student-teacher dialogues for argumentation to
17 happen. From an epistemological perspective students need to understand and practice
18 the norms and criteria underlying scientific work (Hogan & Maglenti, 2001; Zohar &
19 Nemet, 2002). We may think the laboratory context automatically also offers this type of
20 teaching, but the “recipe-like” nature of activities presents a misleading epistemology.
21 Teaching, therefore, should break this pattern by actively encouraging and guiding
22 students to propose, justify, evaluate and criticise knowledge claims.
23
24

25
26 Applying Jiménez-Aleixandre’s (2008) analysis of learning environments supporting
27 argumentations the following characteristics of laboratory tasks indicate positive
28 outcomes for argumentation:
29

- 30
- 31 • Students should be faced with and encouraged to discuss alternative theoretical
32 hypotheses and alternative actions (e.g. different designs and ways of
33 investigating hypothesis). Such hypotheses and actions may be presented by the
34 teacher or generated by the students; - the main point is reflecting on and debating
35 alternatives rather than seeing one alternative as the only “correct” one.
36
 - 37 • Students should be encouraged to use data as evidence to back their claims and
38 actions. This use includes examining the quality of the data and presenting
39 limitations to what conclusions can be drawn, and it should be emphasised that
40 uncertainty in science is related to both theories and data.
41
 - 42 • The teacher should play the role of a facilitator which actively engages in the
43 scientific debate. He or she should challenge students and ask for evidence to
44 justify their views (Simon, Erduran & Osborne, 2006); for example by asking
45 “Why do you think that?” and “How do we know it?”
46
 - 47 • The teacher should act as an “able peer” who holds the scientific epistemic criteria
48 and who provides scaffolding to the discussions. In a laboratory context this
49 scaffolding also may be helped by the task, as prompts and written rubrics, but it
50 is an important element in the continuous communication between the teacher and
51 the students. The analogy offered by Sandoval and Reiser (2004) is that of
52 “cognitive apprenticeship” into scientific practice: the teacher includes students to
53 conversation built on scientific epistemic criteria.
54
55
56
57
58
59
60

Peer argumentation in the school science laboratory – exploring effects of task features

This rationale shows teachers play a crucial role in creating scientific argumentative discourse. The aim of the current study, however, was to investigate effects of task formats so the teacher's behaviour was held "constant" by operationalising the two last points listed above for all tasks. Emphasis was on finding and exploring alternative strategies for operationalising Jiménez-Aleixandre's (2008) two first points.

The study

The study started by identifying common investigative tasks in the England and Wales science curriculum (REF needed). Although variation exists, some tasks are used repeatedly in many classrooms, partly due to guidance from assessment authorities (Donnelly et al., 1996) and partly for practical reasons: finding good research or genuinely open-ended problems suitable for student investigations in school contexts is thought to be difficult. Good examples are easily established so tend to reoccur in teaching materials and lessons. As mentioned, typical tasks are to find the effects of one or more independent variables on a single dependent variable. Even if tasks are characterised as "open-ended" students are often given a research problem and follow standard procedures to arrive at pre-determined solutions. Newton et al. (1999) observed such "open investigations" being used in 4 out of 23 randomly observed practical work lessons. Two of these typical tasks were chosen for this study (see below).

Investigations are at the heart of the problem of creating argumentation in the science laboratory: they are intended to introduce students to methods of scientific inquiry, but often fail to do so because of practical and pedagogical constraints (Chinn & Malhotra, 2002a). These activities involve no discussion about the evidence generated, or the way a conclusion has been drawn, since data are "unproblematic" and the answer "obvious". The next step in the study was therefore to reflect on ways in which this pattern could be broken and the tasks changed to include more authentic scientific argumentation. Working from the rationale suggested in the introduction and looking towards research studies promoting argumentation strategies in science lessons more generally (e.g. Osborne, Erduran and Simon, 2004) we identified three possible strategies outlined below.

Use of complex data

The first strategy attempts to stimulate scientific argumentation by use of "complex data"; i.e. by leading students into a situation where data do not give one obvious "correct" conclusion. This strategy was built into a task investigating temperature drop when containers of different materials are filled with hot water (Appendix , Task 1, name for students "Container"). The task was set in the context of investigating transportation of hot liquid in metal tanks with different surfaces. Students were provided with metal containers with three different surfaces: black, white and (shiny) grey metal coloured. Being of the same material the only factor separating the containers was the surface colour - this gives a small difference in temperature drop due to heat radiation. The shiny container will have least radiation and consequently the smallest drop in temperature. All temperature drops are small, leading students to produce measurements in which differences in data collected from the three containers are of similar magnitude as the level of uncertainty in each measurement.

*Peer argumentation in the school science laboratory – exploring effects of task features***Conflicting hypotheses**

The second strategy presents students with conflicting theoretical hypotheses, about which they are likely to disagree. The students are asked to debate which of a set of statements are right or wrong *before* starting the investigation, and then to investigate these in experiments. The strategy was implemented into another investigative task common in science lessons: dissolving salt (sodium chloride) in water of different temperatures (Appendix, Task 2). Statements were based on the misconception commonly found among school students that sugar or salt loses mass when dissolved in water (Ebenezer & Erickson, 1996).

Post-investigation discussion

The third strategy differs from the two first in that a discussion is arranged *after* the practical investigation has been finished. Handling equipment and gathering data are complicated tasks so arranging a discussion afterwards is as an alternative that may help students be more reflective about scientific inquiry. A task was developed that followed up the practical investigation in Task 1 (complex data). Students were presented with data related to the investigation collected by imaginary student groups who had used different methods (Appendix, Task 3). Questions in the task were directed at making students evaluate the methods and the data produced, and to debate evidence for the conclusion. Students started by answering questions individually and then debating their answers in order to try reach a common conclusion.

The study investigated the effects of each strategy on 12 -14 year old students' argumentation when working in small groups. The research questions investigated were:-

1. To what extent do these strategies stimulate scientific argumentation?
2. If argumentation occurs, what forms of debate arise for each of these strategies?

Design and methodology

The design was based on testing the three different strategies in ordinary whole class laboratory teaching. The tasks were planned for lessons lasting 60 minutes (a time which is common in secondary schools in England and Wales), at an independent school (i.e. non-state funded) in the North East of England. The participants were three Year 8 (12-14 year old) mixed ability classes, each with 22-23 students. The science teacher (author 2 of the paper) taught all classes. This ensured consistency across the teaching situations and a classroom climate that encouraged argumentation as indicated above.

Two classes carried out all three tasks in small groups. From these classes, four groups, each of three students, were selected for in-depth analysis. From the third class two groups were selected. These acted as "controls" by doing only Task 3, the "post-investigation" that is without first having conducted the matching practical task (1). Groups were sampled to mix gender and ability. Table 1 shows the data gathering design. The design made it possible to compare different student groups doing the same activity to see how a strategy worked for four different student groups. We were also able to compare the same group doing three different tasks to see how a student group reacted to three different strategies. The two groups doing only Task 3 enabled comparison with

Peer argumentation in the school science laboratory – exploring effects of task features

students who did this task with and without carrying out the practical task (Task 1) first. Data gathering was carried out over a period of two weeks.

(Table 1 in about here)

Data gathering was made with video recording, using one high quality Sony DVCAM camera on each focus group, and wireless “fly” microphones attached to the students. This made it possible to record in detail what students did and the conversation between the students in spite of being in a classroom with background noise. The microphones also recorded conversations between focus group students and the teacher. Using video camera and microphones, of course, will influence students’ talk, but students in all groups kept talking freely and their behaviour suggested no obvious constraints due to this situation.

Students produced written reports, one each group, from the investigations. These were collected and used as background information to support and validate data from the video recordings. In Task 3 students also produced written responses that were compared with the videos.

The observed lessons all started with 5 – 10 minutes introduction to the task by the teacher. Emphasis was placed on students being explicit about their thinking and the need to make decisions during the investigations. They were also told that conclusions needed support from evidence and that this should be a particular focus. The students were used to working in the laboratory and familiar with investigative tasks. No specific teaching, however, had been provided on argumentation. The introduction and the clearing up afterwards left 30 to 40 minutes for each practical investigation (Task 1 and 2). Task 3 required less time and lasted 10 to 20 minutes, depending on how quickly students handled the task.

Data analysis

Data analysis started with first transcribing all conversations from the video recordings. The transcripts were checked by a second, qualified academic colleague to ensure their correctness. The remaining data analysis was made from the transcripts, but often with a return to the video recordings to have a better understanding of the context or by confronting students’ written reports. Although this is a small scale study, the analysis was both qualitative and quantitative.

Identifying argumentation discourse

The main focus of the analysis was to identify the type of argumentation discourse occurring in the student groups, but for comparison between tasks we also “measured” frequency and quality of argumentation. Several possible analysis methods were considered, as there is no one obvious method (Erduran, 2008). Watson (2004) counted claims made by students and looked at how often these were supported by data. Newton et al. (1999) focused more generally on the orientation of students’ work, e.g. “group discussion” and “closed experiments”, and measured time devoted to each of these. Students’ discussion in the different orientations were then studied qualitatively, and

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

together with the quantitative time measure gave an indication of the amount of argumentation happening. A third approach, adopted here, is to identify “argumentation units” using Toulmin’s (1958) Argumentation Pattern (TAP). TAP describes argumentation with a claim being supported by data, and with various ways of strengthening or undermining this relationship. Each unit, of course, can last for various lengths of time, but this is subordinate to the quality of the argumentation contained. The argumentation measure is therefore given by number of units (more units indicate more frequent argumentative discourse) and the quality of each. Counting and scoring the units is not straightforward, because of the difficulty related to deciding what is claim, data and warrant (Erduran, 2008). When a new claim is made, for example, it may sometimes be judge as a simple “counter claim” (meeting a claim with another claim), and therefore belong to the same unit, or be seen as the start of a new unit. A depending factor is how it is supported with data and warrants. The numbers produced for units should therefore be read with some caution. They give meaning mainly when compared within the study: here, the researchers have been able to compare units and agree on definitions used consistently in the three tasks. In accordance with Zohar and Nemet (2002), claims made without any justification or not met with a counter claim were not recognised as an argumentation unit.

Determining the quality of argumentation discourse

The criteria used for scoring quality of argumentation units relates to low quality argumentation being “sparse” with few backing or rebutting elements and high quality argument “rich” in such elements. To score this issue we used a classification system developed by Erduran, Simon and Osborne (2004, see Table 2). Other frameworks were considered (an overview is presented in Sampson & Clark, 2008), but Erduran et al.’s (op cit) framework was selected because the five level scale provides a means of grading students’ comments. The key discerning factor is the presence or absence of rebuttals. In the two lowest levels there is no questioning of claims. Students either meet a claim with another claim (Level 1) or use some form of argument to support their claim (Level 2). The three next levels considered the quality and quantity of rebuttals. Identifying and coding argumentation units were equally important for the qualitative analysis of the group discussions as it was for the “scoring” and “counting”. The close inspection of units served as a means for understanding the nature of the laboratory discourse.

(Table 2 in about here)

Two researchers identified units and carried out the coding for each piece of transcript. When unites were agreed, inter-coder reliability was 70 - 80 % when coding their levels. Disparities were resolved through discussions.

Student orientation

An additional coding of students’ “orientation”, or focus, while working on the tasks was also conducted. This coding was not planned in advance but derived from the data as an attempt to characterise ways in which students solved the investigation tasks. Theoretically the coding has support in Klahr, Fay and Dunbar’s (1993) *Scientific Discovery as Dual Search* (SDDS) model. This model suggests that someone (students and scientists) working on an inquiry task may operate in different “problem spaces”. The

Peer argumentation in the school science laboratory – exploring effects of task features

model has two problem spaces, the *experimentation space* and the *hypothesis space*, but Klahr et al. add a third dimension, making in total three orientations:

- a) Experimentation: Students are focused on data gathering and handling of equipment
- b) Hypothesising: Students are focused on explaining the observed phenomena by use of scientific theories and concepts
- c) Co-ordination and evaluation: Students are focused on co-ordination and evaluation of evidence to draw a conclusion.

By using the SDDS model as a theoretical underpinning is suggested that students may be operating in these orientations disjointedly and with some problems combining them or moving from one orientation to another. Further details about how the coding was conducted will be given in the result section below.

Results

We will start by presenting two examples demonstrating students' complete work on Task 1 and 2. The examples illustrate characteristics of group work being "rich" (No 1) and "poor" (No.2) on argumentation. As such, these represent extremes in the findings. Number 1 is Group A carrying out Task 1 (Container, complex data) and Number 2 is Group C carrying out task 2 (Solution, conflicting hypotheses). Students in the groups have been given fictitious names with first letters matching the group letter. Student comments are reported verbatim with editorial additions shown in square brackets to ensure the context is clear.

Example 1: Group A on Task 1, "Container" with complex data

After having been informed about the task and organised as a group the three girls in Group A went directly on to collecting equipment and setting up the experiment. No discussion occurred about the purpose of the task.. They exchanged single comments as they went along with the preparations, repeating information from the task sheet or given by the teacher, to clarify and agree what they were doing. For example (brackets show time in minutes: seconds):

Ann (0:25): We need to explain clearly what goes into the container.

Nearly nine minutes later, three metal containers with different surfaces (white, black and metal), each with a thermometer were lined up at the table and filled with hot water from a kettle. Then came explicit comments about the research question:

Ali (8:50): We need to do something..

Ann (9:03): We need to see which one ... like keeps the most heat..

It was clear that the students' focus so far had been *what* to do rather than *why*. Having made their first recordings of temperature in each container the discussion continued:

Ann (9:22): Do they [the company mentioned in the task] want to keep it

Peer argumentation in the school science laboratory – exploring effects of task features

hot .. or do they want to keep it cold?

Ada (9:29): I don't really know

Ali (9:29): They want to keep it warm

Ann (9:30): They want to keep it hot, okay

Ann (9:31): Okay, so we see which stays the hottest

Ali (9:32): This is the hottest at the moment...

This was their only explicit “planning” and in the following twelve minutes the students were fully engaged with recording data. As they went along, however, three more methodological issues (italicised) were brought up, all in the same accidental way and each was solved in a single sentence:

First, *how to present data*:

Ali (11:31): So.. we like, make a table

Ada (11:31): [Grunts] Yeah

Second, *what was the temperature of the water when they started?* (They forgot to record temperatures immediately after having poured the water.)

Ada (13:15): What was the temperature at zero Ali?

Ali (13:17): Temperature at zero was 100 degrees for everything, because it had all come out of the ..eh.. kettle.

Third, *how long should they keep taking measurements?*

Ali (21:15): How many minutes do we have to do it for?

Ann (21:17): I don't know

Ali (21:20): Just keep going until she (the teacher) stops us

Ann (21:24): Yeah

During the data gathering analysis was also carried out in a similar way, as the girls started to discuss which container best keeps the temperature. This started with Ali making a comment after the ninth set of readings, in which the temperature in the metal coloured container (“silver”) was 86 °C, the black was 84 °C and the white was 81 °C:

Ali (18:47): I think silver is going to be the best, don't you

Ada (18:49): Yeah

A few minutes later Ali declared the “winning container” and told the teacher they had finished the investigation. The teacher wanted the students to evaluate the data before drawing a conclusion, but the students did not see a need for this. When the teacher asked the students *why* the metal surfaced container had best kept the temperature, it became clear that the students had expected the black container “to win”:

Ada (24:58): [Be]cause it absorbs heat, like when the Sun is...

Ali (24:58): Like when you wear black clothes in the Sun

Peer argumentation in the school science laboratory – exploring effects of task features

Ada (24:58): Yeah, in the Sun and then it absorbs the heat

These ideas, however, were never discussed among the students while doing the investigation, or in the students' report, which just presents the measurements and confirms that the metal coloured container came out with the smallest fall in temperature.

Example 2: Group C, Task 2: "Dissolving" with conflicting hypotheses

The students in Group C, two boys and a girl, started by reading out the three statements in the task (see Appendix 2) and discussing which were "right" or "wrong". Although lasting less than two minutes the discussion made a basis for the further investigation. Two students (the boys Cameron and Callum) supported the statement "mass cannot disappear" and claimed the opposite statement "there is always a small loss of mass" therefore had to be wrong. The third student, Cynthia, was less sure and thought statement B might be right. Statement C (mass depends on the temperature of the water) was declared wrong by Callum. The two other students expressed some doubts about this, but Callum's self-assuredness silenced their views. The final outcome, decided mainly by Callum, was to do experiments to test statements A and B.

The students decided on using 25ml water and 1 g salt to test hypothesis A. These were measured independently before the salt was dissolved in the water and a new measurement made of the solution. With small amounts, the measurements were not exactly the same before and after dissolving the salt. There seemed to have been a loss of mass, which confirmed Cynthia's view:

Cameron (16:15:): Some is gone... B is right, I think

Cynthia (16:16): I said B. Didn't I say B!

At first they were all convinced about this conclusion, even Callum:

Callum (16:32): So there is always a small loss of mass when it dissolves...
The mass will increase, but less than the mass of the salt added.

Cameron (16:47): That is right.

Callum 16:47): So B is right.

Cynthia then suggested they had *disproved* statement A only and that they still had to *prove* statement B. A 15 minute discussion started in which Callum tried to convince the others that having a test disproving A is the same as proving B to be right. The outcome of the discussion was an agreement that they should do the same test again. When doing this they tried to copy the exact amounts used in the first experiment, arguing that "things had to be kept the same". This, however, was difficult and their next set of readings was different:

Callum (21:26): Shouldn't it be the same as that [compares measurements from the two experiments]

Cynthia (21:30): No, it wasn't, like, precise. You can't get ... or you can, but it would be really hard to get exactly the same amount

Peer argumentation in the school science laboratory – exploring effects of task features

This comment from Cynthia gave Callum a sudden understanding that hypothesis A still might be right and that the unexpected results in the first experiment was due to measurement error:

Callum (21:36): No, wait a moment (eager). That means that that [statement A] could be right. Because, it is really hard to get the exact measurements.

He continued:

Callum (21:56): I do not think we lose anything. It is just impossible to make it precise.

Callum tried to convince the others about this. They, however, still thought they should stay with the outcome of the measurements, which indicated a loss of mass (i.e. supported statement B). The discussion included detailed analysis of the experiment and why or why not the final outcome should be trusted. Callum did not manage to convince the others on his view.

Patterns of orientation in students' investigations

The two examples show different approaches to the investigations. Example 1 demonstrates an algorithmic approach, focusing on data gathering to answer a question. The girls implicitly “knew” what to do, not needing planning, and they unreservedly trusted the outcome of the measurement, not perceiving a need to evaluate. A conclusion could be drawn, and the investigation was “finished” as soon as the last measurement was made. Example 2 developed differently. More time was spent on planning and on conceptual discussion, but the most characteristic difference from Example 1 is extensive discussion about what conclusion could be drawn from the data. This was absent in the first group, but filled more than half the time of the second group. The difference has an obvious importance for occurrence of scientific argumentation.

As a way of demonstrating these differences across all groups and tasks a decision was made to code time spent on the different orientations of students' discourse. Social talk, talk about the task (rather than problems in the task) and discussing/writing the report were excluded from this analysis and coded as “other”, as these were not our focus. The transcript was the main source of coding but with support from the video pictures. Sequences of the group investigation were given a code a) Experimentation, b) Hypothesising or c) Co-ordination and evaluation, and the time spent on each of these was added up.

Figure 1 shows how time was spent by the groups on each of the orientations over the tasks they carried out, as shown in Table 1. The category “other” is excluded and the percentage is therefore a relative distribution of time spent between the three main categories.

[INSERT Figure 1 about here]

Peer argumentation in the school science laboratory – exploring effects of task features

The graphs show that Task 1, Container with complex data, prompted students to focus on “experimentation”. Groups A (the first example presented above), held this focus 90% of the time, hardly devoting any time to “hypothesising” and “coordination and evaluation”. Group C on Task 2, Solution with conflicting hypotheses, (the second example presented above) is an exception, spending more time on coordination and evaluation than on experimentation. Task 2 stimulated more discussion about coordination and evaluation than Task 1. However, we sense some “randomness” in the data: how a task develops for the groups is inconsistent. Groups B and D exhibit similar patterns for Tasks 1 and 2, indicating that orientation is group-related (groups having the same orientation across tasks), while the opposite is the case for Groups A and C, indicating that orientation is task-related. All in all, this indicates a situated effect, suggesting that the way in which a task develops for a pupil group is influenced by factors happening during that particular event.

Task 3, the post-investigation discussion, naturally has a higher percentage of the time on the two last categories, since it was non-practical. We, see that students, except for Group D, balance their time between “hypothesising” and “coordination and evaluation”. This indicates that hypothesising, which was obviously very difficult to stimulate during the practical work, was more easily stimulated in this task format.

Argumentation and students’ orientation

Table 3 summarises the number of argumentation units in each of the three tasks. We see that Task 2, Container, had more units than the two other tasks, generated by groups A and C. Task 3, the post-investigation discussion of Container, had fewest units. It is, however, important to keep in mind that Task 2 lasted 30-40 minutes while Task 3 was solved within 10 to 20 minutes

(Table 3 about here)

Figure 2 presents the level of the argumentation units in Table 3 when scored against Erduran et al.’s (2004) framework. We judged most argumentation at level 2, in which students present a claim with some form of justification, but without rebuttals. Tasks 2 and 3 have relatively more units at higher levels than Task 1.

(Figure 2 in about here)

Figure 3 presents the number of argumentation units in each of the three “orientations” presented earlier, but summed up across all tasks. 12 % only (10 out of 81) happened in “experimentation”, which took on average more than eighty percent of students’ time in the practical tasks. Argumentation occurred most frequently when students were coordinating and evaluating the evidence and conclusions.

(Figure 3 in about here)

Peer argumentation in the school science laboratory – exploring effects of task features

Figure 4 suggests that argumentation unit quality was enhanced to 2.7 on average when students were “hypothesising” and doing “coordination and evaluation”, compared to 1.8 for “experimenting”. Argumentation in the “experimentation” mode meant mainly students making claims in relation to the data recording. For example, in Task 1, claiming one container to be “best” and supporting this with the measurements of temperature. A claim like this was unlikely to be rebutted by another student and the argument unit was very short. Some discussion occurred about methods and triggered some more advanced structures, but this was rare and also relative short. For example, Darren in Group D made the claim that their experiment on dissolving salt (Task 2) was “false“:

Darren (12:03): ..we have changed the amount of water since the first time we weighed it, so this mass will be totally different. So this experiment is false.

Some discussion developed from this, but the problem was solved by carrying out the experiment again. The argumentation occurring when “hypothesising” had higher levels because of students rebutting each others explanations, but this again was not a very frequent event. In the “coordination and evaluation” mode the type of argumentation had much more variation. Students could go into detailed discussion about data in order to solve disputes about what conclusions to draw. This has already been demonstrated by the examples given in the Group C solving Task 2, but a more detailed illustration of the type of argument will be demonstrated. This is Group D discussing the statement that “mass is dependent on temperature of the water” in the same task. Darren made the claim that temperature has no effect, but was challenged with the rebuttal that “salt was at first observed at the bottom of the beaker, but disappeared when the temperature increased”; based on the warrant that solid salt and dissolved salt do now weigh the same. Darren first agrees with the observation but attacks the warrant:

Darren (24:00): The higher the temperature the more salt can dissolve, but it does not affect the mass; unless you put more salt in. If you have the same amount of salt and the same amount of water, heating doesn't actually change the actual mass that is in there

He then strengthen is own argument with a rebuttal

Darren (24:25): Unless the water is evaporating.

Although the values in Figure 2 to 4 should be read cautiously, considering the subjective character of identifying and scoring argumentation units, the results have a clear trend: the data gathering part of laboratory tasks does not invite argumentation discourse. Such discourse happens mainly when students try explaining observations and reflecting on the evidence for their conclusions.

(Figure 4 in about here)

*Peer argumentation in the school science laboratory – exploring effects of task features**What effects on argumentation relate to the strategies implemented in the tasks?**Complex data: Task 1, Container*

Considering that the temperature differences identified in the groups were 1 - 3 °C and that students sometimes disagreed when reading the thermometer, there were good grounds for discussion about the evidence for deciding which container had smallest heat loss. However, all groups concluded that “silver was best” regardless of how small the temperature differences were. When challenged by the teacher students did point towards “accuracy” of their data gathering. Their response reveals that “accuracy” was not interpreted in terms of *measurement error*, but rather as *measurement strategy*. They defended their conclusion with comments about measurement frequency and regularity, but never doubted that a single measurement could be wrong or uncertain. Hence, complex or “uncertain” data in itself did not create any discussion: there was no need to discuss evidence or the conclusion since measurements gave “the correct answer”.

Conflicting hypotheses: Task 2, Dissolving Salt

This strategy prompted initial discussion, bringing forward conflicting views and predictions. The students, however, soon put the conceptual discussion aside and focused on data gathering. Figure 1 shows a small proportion of time spent on “hypothesising” relative to “experimentation”. Data gathering took most time and in this “mode” little attention was paid towards other parts of the investigation. Figure 1 also shows that more time in this task was spent on “co-ordination and evaluation” compared to Task 1, due to conflicts occurring when data did not give the predicted answer. Students supporting statement A (no loss of mass) expected measurements before and after dissolving the salt to be *exactly* the same, without considering differences due to measurement error, and this problem had to be solved. Several outcomes occurred, which aligns with the findings in Chinn and Malhotra (2002b):

- Group A denied the problem and simply accepted their measurements without saying whether they supported or refuted the hypothesis. That is, the data became the final answer to the task.
- Group B rejected their original hypothesis and concluded that salt loses mass when being dissolved in water, thus accepting the data as “true”;
- Group C at first rejected their original hypothesis, but then discussed the evidence;
- Group D ignored anomalous data and kept the original hypothesis, thus accepting the hypothesis as “true” over the data.

Each of these situations gave a different ground for argumentation. If, as in group B, students believed in the data, argumentation was short:

Ben (24:15): So we were wrong.

Bob (24:29): 183.69 [reads out the mass of the dissolved salt in solution]

Ben (24:33): Did you say it should be exactly....?

Bea (24:40): We might have missed out a bit of salt

Bob (24:55): So we might lose a bit of mass, so the hypothesis might be right

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Bob (24:58): Our hypothesis was incorrect!

Bea (25:05): So it will increase a little bit, but not as much as

Bea and Bob here raise a rebuttal, but these carry little weight because data are “correct” and give the “final answer”. Group D held similar certainties about the hypothesis, ignoring the data because they *knew* their hypothesis was correct. Only Group C, who accepted data as uncertain, argued with higher frequency and advanced into more use of rebuttals. Uncertainty in the data opens up the possibility that a hypothesis might be seen as correct or not, depending on data quality.

Two conclusions arise from the conflicting hypotheses task. Firstly, presenting alternative hypotheses stimulates argumentation. This happened especially in the initial phase when students tried to resolve conflicts between hypotheses and in the experimental phase when discussing matches between hypotheses and data. Such conflicts also occurred in Task 1, where students observed the metal-surfaced container best keeping the temperature despite believing the black container was “correct”, but this task did not initiate the same amount of argumentation. We attribute the difference to the explicitly-stated hypotheses in Task 2 that prompted initial discussion. Secondly, the frequency and quality of argumentation is strongly influenced by students’ understanding of uncertainty in the data. Students’ belief in the “truth” of data limits the quantity and quality of argumentation.

Post-investigation discussion: Task 3

This format prompted a focused and effective discussion generating more argumentation units per unit time about the “container” investigation data. The outside laboratory setting enabled students’ attention to be readily directed towards reflecting on both data and method. Surprisingly, students also paid more attention to “hypothesising”, although this was not explicitly required. Figure 1 shows almost all groups used 50% or more of the time explaining *why* a metal-surfaced container should keep temperature better than the white and black ones. This matter was neglected by students doing Task 1. The alternative hypotheses discussion stimulated argumentation with frequent use of warrants and rebuttals. Group E, for example, suggest first that the explanation lies in the property of the metal, but raised the rebuttal that all containers were the same type of metal. They pointed out that black is known to “attract” heat. Next they suggest “layers of the paint” as an explanation, but again found a rebuttal by consulting the data:

Eric (21:27): black might have had more paint on, but metal still does better.

Their last suggestion was that the temperature differences may be caused by “different physical properties of the paint”, to which they did not manage to form any rebuttal and therefore kept as their final explanation.

This task also showed that doing the related practical task (Task 1) first did not generate more or “better” argumentation (see Table 1). In fact, some group discussions indicate the opposite. Groups A to D, who did Task 1 first, compared the fictitious Task 3 data with their own investigation and results, so did not see a need for discussion: their own method, measuring temperature every minute, was judged the “best” strategy and their

Peer argumentation in the school science laboratory – exploring effects of task features

own data offered the “correct” conclusion. Groups E and F, who had not done Task 1 beforehand, discussed the four sets of data provided, drew a conclusion and used methodological criteria to decide which was experiment was best. These observations indicate that students collecting their own data may *restrict* rather than stimulate scientific argumentation.

Discussion

Changing laboratory teaching from a “positivist” tradition (Driver et al., 2000), towards nurturing authentic scientific inquiry was the motivation for this study. To date, scientific argumentation, which should be a natural part of any inquiry process, has played a minor role in laboratory teaching. The oversimplified methods implemented in many laboratory tasks guide students from research problems to final conclusions without the need to raise questions about the method used or the quality of evidence collected. Consequently, data gathering becomes the main focus for student activity, while other elements of the inquiry process are neglected (Abrahams & Millar, 2008; Newton et al., 1999; Watson, 2004). Extant research has also shown that school experiments, for the above reason, teach students a misleading picture of scientific inquiry, reinforcing an unscientific epistemology by encouraging the belief that science is a simple, algorithmic form of reasoning (Chinn & Malhotra, 2002a; Germann, Haskins, & Auls, 1996). Our study challenges this practice by testing the effects of three different strategies in investigative tasks.

The findings point to difficulties in changing existing laboratory teaching practices. Despite working in a learning environment that strongly encouraged evaluation and discussion of the quality of the evidence, and the task strategies, more than eighty percent of students’ time and attention were focused on data gathering. Our evidence shows that this “mode of working” is the least stimulating for scientific argumentation. The three strategies diverted students’ attention towards “hypothesising” and “coordination and evaluation” with mixed success. Students ignored complex data (Strategy 1) in Task 1 (Container) identifying the “best” container, without thinking of the possibility that their data may not fit this purpose. Conflicting hypotheses (Task 2, Dissolving Salt) produced more argumentation than Task 1, but did not for all groups. Only one group examined the evidence thoroughly, by accident, because they measured twice with different outcomes. Post-investigation discussion (Task 3) generated sufficient distance from the investigation to enable students to consider evidence in relation to the conclusion. But, even then, some students solved the task as looking for the “correct” approach and the “right” conclusion. Jiménez-Aleixandre et al. (2000) suggest that hypothetical data provided by the teacher generate different patterns of discussion than empirical, uncertain data, gathered by students themselves. A conclusion from this study is that students’ working from their own data does not guarantee their engaging in evaluative debate.

The study points towards two key issues in understanding why it is so difficult to engage students in scientific argumentation while working in the laboratory. First, it support Kelly’s (2005) claim that science inquiry is an epistemic practice which requires understanding of the methodological components and criteria involved. Our findings reveal how students take data to be “true”, with no concept of “uncertainty”. We were surprised by the extent to which students put aside personal beliefs to accept a

Peer argumentation in the school science laboratory – exploring effects of task features

measurement uncritically. Even when admitting an error, the conditions under which the measurement was made rather than the measurement itself were responsible - a measurement was always right. Watson (2004:34) similarly observed:

The students carried out their tests and accepted the results of their tests as proof: the results justified claims in an unproblematic way.

This has tremendous implications for any debate regarding evaluation and coordination of evidence: if data are “true”, there is no need for argumentation. Interestingly, at times students could imply they understood that data had uncertainty, but their interpretation of “uncertainty” did not focus on measurement error, but on procedures for data gathering, such as measuring at regular intervals. On one hand, this demonstrates a failure of laboratory teaching: students learn procedures mechanistically, using them without epistemological understanding. Early laboratory teaching should focus on epistemological understanding and less on experimental procedures. On the other hand, it also demonstrates that the strategies trialled in this project were not sufficient to change students’ understanding of the epistemic nature of scientific inquiry.

Second, of significant importance for laboratory-based scientific argumentation, is the restraint relating to “working modes”. Klahr et al.’s (1993) SDDS model offers theoretical support for this. Their “problem space” concept suggests scientists operate in three separate modes, “experimenting”, “hypothesising” and “coordination and evaluation”, where the last mode requires a combination of the two first. This model had strong support in our data. Empirically, however, it was surprising to find how strictly students kept to one of the modes and how difficult they found it to change from one to another. Students spent most time in “experimentation” mode, which may seem natural as they are working with equipment and asked to do data gathering. “Hypothesising” could have been a natural mode, but was rarely found while students were working with the equipment. Only when they put the equipment aside and sat down to discuss the investigation, as in Task 3 was this mode apparent. The reason for this, we believe, is that mental effort is required to leave one mode for another. Operating in a mode means seeing the task in a particular way, being aware of possibilities and restrictions for what is relevant and important. To enter another mode involves changing the conception of the task, so needs some stimuli. This applies when going from “experimentation” to “hypothesising”, but is even more relevant for “coordination and evaluation”. While the two first modes are familiar to students, this last mode is relatively unknown. Many students looked rather puzzled when being asked to evaluate their conclusions and lacked a strategy for this. Familiarising students with the mode of coordination and evaluation may be a start towards establishing argumentation in the laboratory. That means establishing some understanding of the *purpose* of the task and demonstrating some examples for strategies to handle the problems involved.

Although argumentation in a laboratory context has unique features, similarity is found with research conducted in other contexts. Roth and Roychoudhury (1992), De Vries, Lund and Baker (2002), Sherman and Klein (1995) and Zohar and Nemet (2002) suggest that a structured approach is important in generating productive discussion. Students need help when putting forward and identifying claims and evaluating using scientific

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

knowledge and data. Other research suggests that establishing argumentation structures in teaching takes time, and over two weeks only limited improvement may be expected. Osborne, Erduran and Simon (2004) report their nine month intervention to be too short for students to develop sufficient skills and abilities. This is supported by Zoller, Ben-Chaim, Pentimalli and Borsese (2000), investigating first year college undergraduates development of critical thinking. They point towards the need of recurrent opportunities to engage in same type of activities. Only when students have obtained some understanding of the pattern and purpose of activities may we expect the teaching to become efficient. The message, however, from these and other studies is that teaching of argumentation *is possible if it is explicitly addressed and taught*. The current study indicates that the same applies to the science laboratory, offering a sought for opportunity to practice argumentation in a scientific, as opposed to the more common socio-scientific, context.

The difficulty of establishing argumentation in the laboratory, of course, should not be underestimated. Laboratory teaching is deeply rooted in the logical-empiricist tradition and has a long way to go before socio-constructivist epistemological perspectives become commonplace. If, as the current research suggests, this is hindered by the psychological constraints of students operating in different “working modes”, we should not be too optimistic for rapid change. More research is needed to establish better understandings of efficient strategies. The present study emphasised “implicit” teaching of epistemological criteria, and a useful alternative to be explored is what happens if these are made more explicit. A longitudinal intervention may provide revealing information. Laboratory work needs to prove less situated with “random incidents” being the main pattern. The hope is that recurrent use of task strategies may establish structures and understanding that students carry with them from one activity to the next. A last suggestion is the need for wider use of task strategies. Knowing that a teacher’s ability to foster a context of argumentation is crucial (Osborne et al., 2004), the current research, with teaching being conducted by one of the researchers, has clear limitations. The strategies should therefore be tested with other teachers and across different educational cultures. This and other research, however, should be given priority as the laboratory is too important as an arena for training scientific argumentation to be left unattended.

References

- Abrahams, I., & Millar, R. (2008). Does Practical Work Really Work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1464-5289.
- Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honour of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum.
- Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings *The Journal of the Learning Sciences*, 2(2), 141-178.
- Chinn, C. A., & Malhotra, B. A. (2002b). Children’s Responses to Anomalous Scientific Data: How Is Conceptual Change Impeded? *Journal of Educational Psychology*, 94(2), 327-343.

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3 Chinn, C. A., & Malhotra, B. A. (2002a). Epistemologically Authentic Inquiry in
4 Schools: A theoretical Framework for Evaluation Inquiry Tasks. *Science*
5 *Education*, 86, 175-218.
- 6 De Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue:
7 Explanation and argumentation as vehicles for understanding scientific notions.
8 *Journal of the Learning Sciences*, 11, 63–103.
- 9
10 Donnelley, J., Buchan, A., Jenkins, E., Laws, P., & Welford, G. (1996). Investigation by
11 Order. Policy, curriculum and science teachers' work under the Education Reform
12 Act. Nafferton: Studies in Education.
- 13 Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific
14 argumentation in classrooms. *Science Education*, 84(3), 287-312.
- 15 Ebenezer, J., & Erickson, G. (1996). Chemistry Students' conceptions of Solubility: A
16 Phenomenography. . *Science Education*, 80(2), 181-201.
- 17 Erduran, S. (2008). Methodological Foundations in the Study of Argumentation in
18 Science Classrooms. In S. Erduran & M. Jimenez-Aleixandre (Eds.),
19 *Argumentation in Science Education. Perspectives from Classroom-Based*
20 *Research*. Dordrecht: Springer
- 21 Erduran, S., Simon, S., & Osborne, J. (2004). Tapping into Argumentation:
22 Developments in the Application of Toulmin's Argument Pattern for Studying
23 Science Discourse. *Science Education*, 88, 915-933.
- 24 Ford, M. (2008). Disciplinary Authority and Accountability in Scientific Practice and
25 Learning. *Science Education*, 92, 404 – 423.
- 26 Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology
27 laboratory manuals: Promoting scientific inquiry. *Journal of Research in Science*
28 *Teaching*. *Journal of Research in Science Teaching*, 33(5), 475-499.
- 29 Gott, R., & Duggan, S. (2003). Understanding and using scientific evidence. How to
30 critically evaluate data. London, UK: SAGE Publications.
- 31 Gott, R., & Duggan, S. (2007). A framework for practical work in science and scientific
32 literacy through argumentation. *Research in Science and Technological*
33 *Education*, 25(3), 271 - 291.
- 34 Herrenkol, L. R., Palincsar, A. S., DeWater, L. S., & Kawasaki, K. (1999). Developing
35 scientific communities in classrooms: A sociocognitive approach. *The Journal of*
36 *the Learning Sciences*, 8(3 & 4), 451-493.
- 37 Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: foundation
38 for the 21st century. *Science Education*, 88, 28-54.
- 39 Hofstein, A., Kipnis, M., & Kind, P. M. (2008). Learning in and from science
40 laboratories: enhancing students meta-cognition and argumentation skills. In C. L.
41 Petroselli (Ed.), *Science Education Issues and Developments* (pp. 59-94). New
42 York: Nova Science Pub.
- 43 Hogan, K., & Maglienti, M. (2001). Comparing epistemological underpinnings of
44 students' and scientists' reasoning about conclusions. *Journal of Research in*
45 *Science Teaching*, 38, 663-687.
- 46 Jimenez-Aleixandre, M. (2008). Designing Argumentation Learning Environments. In S.
47 Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education*
48 *Perspectives from Classroom-based Research*. Dordrecht: Springer.
- 49 Jimenez-Aleixandre, M., & Erduran, S. (2008). Argumentation in Science Education: An
50 Overview. In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in*
51
52
53
54
55
56
57
58
59
60

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2 Science Education. Perspectives from Classroom-based Research. Dordrecht:
3 Springer.
4
- 5 Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the lesson"
6 or "Doing Science": Argument in High School Genetics. *Science Education*, 84,
7 757-792.
8
- 9 Katchevich, D., Mamlok-Naaman, R., & Hofstein, A. (2010). Argumentation in the
10 Chemistry Laboratory: Inquiry and Confirmatory Experiments, NARST Annual
11 Conference. PA: Philadelphia.
- 12 Kelly, G. (2005). Inquiry, Activity, and Epistemic Practice, NSF Inquiry Conference
13 Proceedings <http://www.ruf.rice.edu/~rgrandy/NSFConSched.html> . , Inquiry
14 Conference on Developing a Consensus Research Agenda Rutgers University.
- 15 Kim, H., & Song, J. (2006). The Features of Peer Argumentation in Middle School
16 Students' Scientific Inquiry. *Research in Science Education*, 36(3), 211-233.
- 17 Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A
18 developmental study. *Cognitive Psychology*, 24(1), 111-146. *Cognitive*
19 *Psychology*, 24(1), 111-146.
- 20 Kolstø, S. D., & Mestad, I. (2005). Learning about the nature of scientific knowledge: the
21 imitating-science project. In K. Boersma, M. Goedhart, O. D. Jong & H. Eijkelhof
22 (Eds.). *Research and the Quality of Science Education*, 247-258. Dordrecht:
23 Springer.
- 24 Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*.
25 Cambridge, MA: MIT Press.
- 26 Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: Univ. of Chicago
27 Press.
- 28 Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the
29 pedagogy of school science. *International Journal of Science Education*, 21(5),
30 553-576.
- 31 Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in
32 school science. *Journal of Research in Science Teaching*, 41(10), 994-1020.
- 33 Roth, W. -M., & Roychoudhury, A. (1992). The social construction of scientific concepts
34 or the concept map as conscription device and tool for social thinking in high
35 school science. *Science and Education*, 76, 531-557.
- 36 Sampson, V., & Clark, D. B. (2008). Assessment of Ways Students Generate Arguments
37 in Science Education: Current Perspectives and Recommendations for Future
38 Directions. *Science Education*, 92(3), 447-472.
- 39 Sadler, T.D., Chambers, F.W. & Zeidler, D.L. (2004). Student conceptualizations of the
40 nature of science in response to a socioscientific issue. *International Journal of*
41 *Science Education*, 26(4), 387-409.
- 42 Sandoval, W. A., & Reiser, B. J. (2004). Explanation driven inquiry: Integrating
43 conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3),
44 345-372.
- 45 Sherman, G. P., & Klein, J. D. (1995). The effects of cued interaction and ability
46 grouping during cooperative computer-based science instruction. *Educational*
47 *Technology Research and Development*, 43, 5-24.
- 48 Simon, S., Erduran, S., & Osborne, J. (2006). Learning to Teach Argumentation:
49 Research and development in the science classroom. *International Journal of*
50 *Science Education*, 28(2-3), 235-260.
- 51
52
53
54
55
56
57
58
59
60

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3 Tapper, J. (1999). Topics and manner of talk in undergraduate practical laboratories
4 International Journal of Science Education, 21(4), 447 – 464
5 Toulmin, S. (1958). The uses of argument. Cambridge: Cambridge University Press.
6 Vygotsky, L. S. (1978). Mind in society: The development of higher psychological
7 processes. Cambridge, MA: Harvard University Press.
8 Watson, J. R. (2004). Students' discussion in practical scientific inquiries. International
9 Journal of Science Education, 26(1), 24-45.
10 White, R. T., & Gunstone, R. F. (1992). Probing Understanding. : Falmer Press.
11 Zeidler, D., Sadler, T., Simmons, M., & Howes, E. (2005). Beyond STS: A research-
12 based framework for socioscientific issues education. Science Education, 89,
13 357–377.
14 Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills
15 through dilemmas in human genetics. Journal of Research in Science Teaching,
16 39, 35–62.
17 Zoller, U., Ben-Chaim, D., Pentimalli, R., & Borsese, A. (2000). The disposition towards
18 critical thinking of high school and university students: An inter-intra Israel-
19 Italian study. International Journal of Educational Research, 22, 571-582.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

*Peer argumentation in the school science laboratory – exploring effects of task features***Appendix 1 Task 1, Container****Container**

A company transporting hot liquid is buying a new tanker lorry and wants to know what colour the tank should be. You are asked to do an experiment to help give them advice if the colour has any effect on loss of heat.

You should find out *if* the colour has any effect and, if so, *what colour* (black, white or metal) best keeps the temperature.



The company wants to be sure about its decision and therefore would like each scientist to tell what he or she thinks about the experiment. You should:

1. Do the experiment to get the data.
2. Discuss how big the differences are between the containers and how sure you can be that the results are correct.
3. Make an advice to the company, telling them if you all agreed or if there were different opinions.

Your answer to the company must be supported by data and you should try to explain the results.

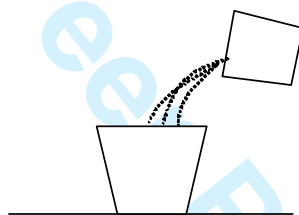
You will have:

- Three types of containers
- Hot water
- Thermometers
- A timer

Appendix 2 Task 2, Dissolving Salt**Dissolving salt**

Some students are investigating what happens to the mass of a container with water when we dissolve salt in it. They have different explanations. With whom do you agree? You should:

1. Discuss each statement and tell **if and why** you think it is right or wrong.
2. Select two statements you are uncertain about or you think are wrong.
3. Then do the experiments to test these two statements.
4. Discuss if the experiments are good enough to draw a conclusion and, if so, why or why not.



Statements:

- a) *When salt dissolves it still weighs the same. The total mass is therefore exactly the same as the water and the salt added.*
- b) *There is always a small loss of mass when something dissolves. The mass will increase but less than the mass of the salt added.*
- c) *The higher the temperature the more salt dissolves, the new mass therefore depends on the temperature of the water.*

You will have:

- Containers
- Hot and cold water
- Thermometers
- A stirrer
- A scale

*Peer argumentation in the school science laboratory – exploring effects of task features***Appendix 3 Task 3 Post-investigation discussion of Container task****Container**

Some students have been doing an experiment to find out if the colour of an object has any effect and on its heat loss, and if so, what colour best keeps the temperature.

They had containers in three different colours: black, white and metal (not painted). In their experiments they filled the containers with hot water and measured the temperature as the containers cooled down. The tables show results from four different groups. They all did the experiment in different ways.

Group A			
150 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
1	90	90	91
2	84	84	85
3	80	79	80
4	76	75	76
5	72	72	73
6	71	70	72
7	70	69	71

Group B			
100 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
5	75	76	78
10	55	57	48
15	40	42	44
20	30	34	36
25	30	27	30
30	23	24	26
35	19	18	19

Group C			
150 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
1	90	90	91
5	72	72	73
16	40	42	44
21	30	31	35

Group D			
100 ml water in each container			
Time (min)	White	Black	Metal
5	75	77	77
10	55	57	58
15	40	42	44
20	29	33	35
30	21	20	22

1. Decide which group you think did the best experiment. Explain why.

Continues

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2. Analyse the results from the group you think did the best experiment and decide which container has least heat loss (best keeps the temperature).

My conclusion is the

Black White Metal coloured They are the same

3. How certain are you about the conclusion? Tick one of the boxes below, or make your own statement.

- I am absolutely certain that the container I identified is the right one.
- The container I have identified is the best in the data, but the group's data gathering is too poor to make a final conclusion.
- The container I have identified is the best in the data, but the results are very uncertain. Another test is likely to show different results.
- It is impossible from the data to conclude if any container best keeps the temperature.
- Other: _____
- _____
- _____

4. Compare your answers to the other students in your group. Decide between you what answers you think are correct.

Our final conclusion is that the _____ container(s) best keep(s) the temperature, because

*Peer argumentation in the school science laboratory – exploring effects of task features***Tables and figures:**

	<i>Task 1: Container</i>	<i>Task 2: Dissolving</i>	<i>Task 3 Discussion of Container task</i>
Group A (Girls)	x	x	x
Group B (Mixed)	x	x	x
Group C (Mixed)	x	x	x
Group D (Boys)	x	x	x
Group E (Mixed)			x
Group F (Boys)			x

Table 1: Design of the study

Level 1	Level 1 argumentation consists of arguments that are a simple claim versus a counter-claim or a claim versus a claim.
Level 2	Level 2 argumentation has arguments consisting of a claim versus a claim with either data, warrants, or backing but do not contain any rebuttals.
Level 3	Level 3 argumentation has arguments with a series of claims or counter-claims with either data, warrants, or backing with the occasional weak rebuttal.
Level 4	Level 4 argumentation shows arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counter-claims.
Level 5	Level 5 argumentation displays an extended argument with more than one rebuttal.

Table 2: Erduran et al.'s (2004: 928) Analytical Framework for Assessing Quality of Argumentation

Peer argumentation in the school science laboratory – exploring effects of task features

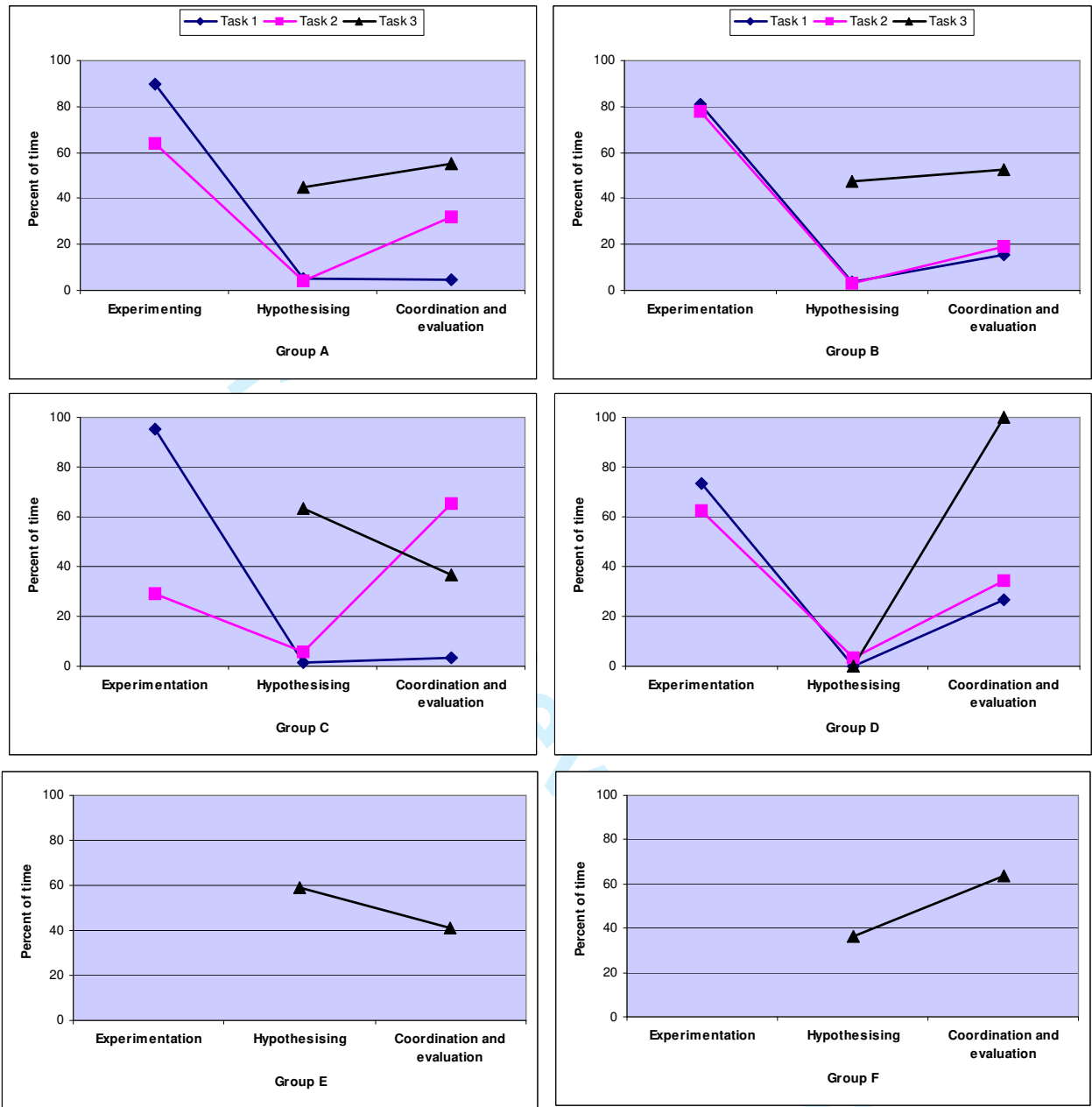


Figure 1: Orientation data: Total time spent by six student groups A-F on experimenting, hypothesising and co-ordinating and evaluating over three different tasks

Peer argumentation in the school science laboratory – exploring effects of task features

	<i>Task 1</i>	<i>Task 2</i>	<i>Task 3</i>	Total
Group A	7	10	8	25
Group B	4	7	6	17
Group C	9	17	3	29
Group D	3	6	1	10
Group E			11	11
Group F			2	2
Sum Gr A to D	23	40	18	81

Table 3: Number of argumentation units identified in each task each group

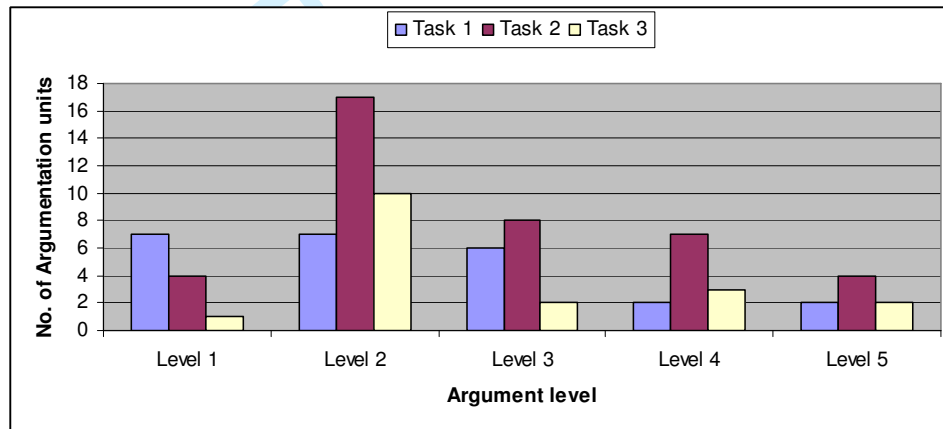


Figure 2: Number and level of argumentation units per task for Groups A to D.

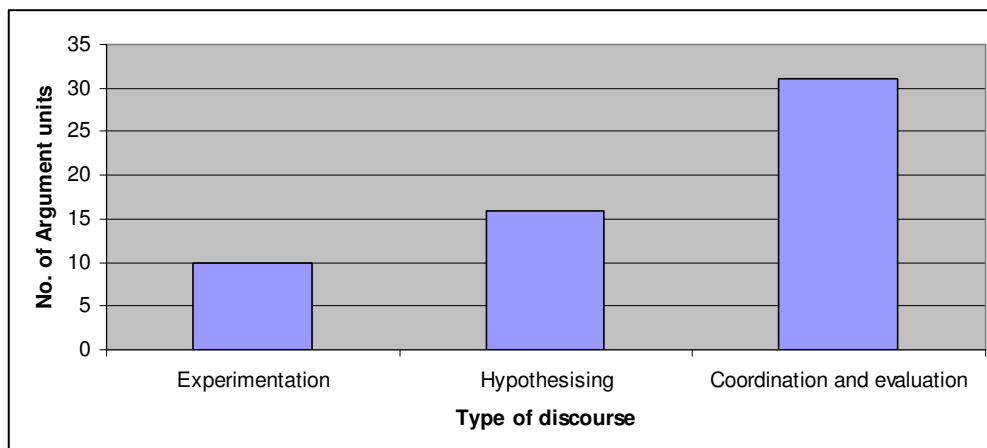
Peer argumentation in the school science laboratory – exploring effects of task features

Figure 3: Number of argumentation units summarised within each type of discourse for groups A to D.

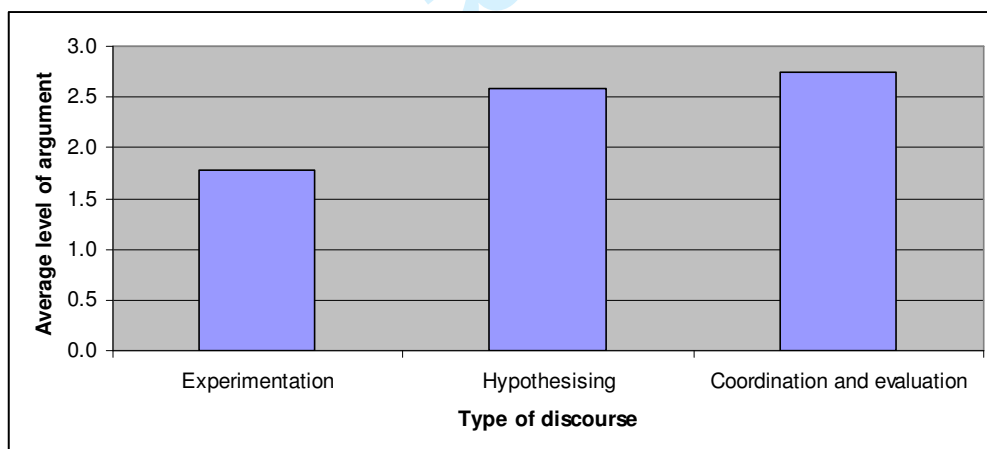


Figure 4: Average level of argumentation within each type of discourse for groups A to D.

**Peer argumentation in the school science laboratory –
exploring effects of task features**

Prepared for International Journal of Science Education

For Peer Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Peer argumentation in the school science laboratory – exploring effects of task features

Abstract

Argumentation is believed to be a significant component of scientific inquiry: introducing these skills into laboratory work may be regarded as a goal for developing practical work in school science. This study explored the impact on the quality of argumentation among 12 – 13 year old students undertaking three different designs of laboratory-based task. The tasks involved students collecting and making sense of complex data; collecting data to address conflicting hypotheses; and, in a paper-based activity, discussing pre-collected data about an experiment. Significant differences in the quality of argumentation prompted by the tasks were apparent. The most argumentation units per unit time were generated by the paper-based task. Where students carried out an experiment, argumentation was often brief, as reliance on their data was paramount. Measurements were given credence by frequency and regularity of collection, while possibilities for error were ignored. These data point to changes to existing practices being required in order to achieve authentic, argumentation-based scientific inquiry in school laboratory work.

Deleted: can

Deleted: of

Deleted: s

Deleted: of three different designs

Deleted: had to

Deleted: y

Deleted:

Deleted: short

Introduction

A rapidly growing interest in *argumentation* among science educators has been fuelled by socio-constructivist theoretical frameworks in science philosophy and learning psychology (Jiménez-Aleixandre & Erduran, 2008). In science philosophy, science theories are no longer seen as products of pure empirical or logical processes but as ideas shaped through critique, debate and revision within the science community (Kuhn, 1962). In learning psychology, learning is seen as originating from socially mediated activities (Vygotsky, 1978). Knowledge and cognitive processes exist in a social milieu and are internalised by the individual through language and active participation. From these key ideas it has been claimed that science education needs to change from a current “positivist” practice, emphasising a misleading picture of science and factual recall of knowledge, towards a practice that helps socialisation of young people into the norms and practices of authentic science (Driver, Newton, & Osborne, 2000). The current paper relates these changes to teaching and learning in the science laboratory. To date, socio-scientific issues have provided a more common focus for argumentation in science education activities, perhaps because preparation for decision-making in political and moral issues students meet in society (Zeidler, Sadler, Simmons, & Howes, 2005) has dominated educational thinking. Laboratory activities may be an attractive alternative drawing the attention towards argumentation in a scientific context and demonstrating something about the architecture of scientific knowledge (Ford, 2008). Osborne, Erduran and Simon (2004) offer a second reason for the dominance of the socio-scientific context, by showing that initiating argument in a scientific context is much harder. To keep argumentation meaningful students need to understand evidence (Koslowski, 1996), which can be challenging for some scientific phenomena. In this view, the laboratory also may be favourable because students familiarise themselves with the phenomenon and generate their own data.

There are, however, obvious reasons that argumentation in the laboratory is difficult (Hofstein, Kipnis, & Kind, 2008). A typical school inquiry more closely resembles the empiricist-logical view of science that dominated science philosophy a century ago than today’s socio-constructivist view (Driver et al., 2000). Scientific method is presented in a step-wise manner leading students steadily along a path from research problem to final conclusion. Thus, they are led in ways that avoid the complex nature of authentic science to ensure that the “right” data appear and “correct” conclusions are made. Open-ended investigations with more degrees of freedom occur, but are limited in their use due to the pedagogical challenges they present to teachers (Hofstein & Lunetta, 2004). Meeting curriculum targets and conforming to assessment practices are also restrictions (Donnelly, Buchan, Jenkins, Laws, & Welford, 1996). Establishing the science laboratory as an efficient place for teaching and learning scientific argumentation requires that we face the challenge of breaking these practices.

Against this background the current paper presents an exploratory study that has investigated the effect of specific pedagogical strategies in laboratory tasks. Chinn and

Deleted: ing of the

Deleted: in order

Deleted: the

Deleted: right

Deleted: being

Deleted: for

Deleted: the

Deleted: R

Deleted: are also made by the need to reach curriculum targets and conform to assessment practices

Deleted: n

Deleted: l

Deleted: B

Deleted: with

Deleted: is a major challenge if we want to establish the science laboratory as an efficient place for teaching and learning scientific argumentation.

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3 Malhotra (2002a) analysed laboratory task formats extensively. They define one group of
4 tasks as “simple experiments”, the main focus in this study. These are typically tasks in
5 which students research the effect(s) of one or two provided variables, with some
6 openness in choice of method. The study is set in the UK context, where such tasks are
7 referred to as investigations. By suggesting particular pedagogical strategies implemented
8 in the tasks, trialling these in “ordinary” teaching contexts and accepting that classroom
9 life is synergetic (Brown, 1992), we hope to shed light on why argumentation in the
10 science laboratory presents challenges and suggest specific actions that may be taken to
11 improve current practices. Before introducing the study the paper reviews research
12 analysing discourse practices in the science laboratory and sets out the underlying
13 rationale.

Argumentation in the school science laboratory

14
15 Science education research literature portrays the laboratory as a complex environment
16 for teaching and learning (Harmon et al., 1997; Hegarty-Hazel, 1990; Hodson, 1993;
17 Lunetta, Hofstein, & Clough, 2007; Séré et al., 1998). On one hand, it seems
18 straightforward, familiarising students to do investigative work. For example, the
19 Assessment for Performance Unit (APU), a large UK-based assessment project, collected
20 data from students doing practical investigations in the early 1980s. Researchers were
21 positively surprised about the relatively advanced levels of student performance (Murphy
22 & Gott, 1984). Despite students having limited experience of carrying out investigations,
23 they approached the APU tasks with ‘confidence and success’ (p 40). Around half of the
24 students were judged to carry out ‘good’ experiments, with only a fifth or less failing to
25 make the expected quantitative measurements. Similar positive data are reflected in the
26 Third International Mathematics and Science Study (TIMSS, Harmon et al., 1997). On
27 the other hand, getting students from a basic level to using the laboratory for more
28 sophisticated learning is difficult. The APU study (op cit) showed surprisingly little
29 progress when 13 and 15 year-old students were compared. Students seem to learn
30 elementary ways of solving laboratory tasks quickly, but do not progress beyond these.
31 One contributory factor is the standardised form of laboratory tasks, which invite students
32 to learn ‘scripts’ (Schank & Abelson, 1977). Many will have observed with Millar et al.
33 (1994) that students start solving investigative tasks without first spending time on
34 planning what to do. The wider meaning, of course, is that the science laboratory has
35 become a place for conducting routine exercises (Lunetta, Hofstein & Clough, 2007) so
36 that students are seldom challenged to reflect on the methods they use and/or
37 observations they make (Hodson, 1996). Recent evidence confirms that this problem still
38 applies to UK school science laboratory teaching. Abrahams and Millar (2008) observed
39 twenty-five randomly selected laboratory lessons and found that teaching *seemed* well-
40 functioning on the surface, as students gathered data and conducted activities as their
41 teacher expected. However, students rarely reflected on their findings and methodology.
42 Unsurprisingly, Newton, Driver and Osborne (1999) document the absence of
43 argumentation in ordinary laboratory teaching. As little as 0.4% of time spent on
44 laboratory tasks was organised as group discussion. In fact, laboratory lessons included
45 *less* discussion than non-laboratory lessons. Teacher questionnaires and twenty-three case
46 studies across Europe suggest a similar trend in other countries (Séré et al., 1998).
47 Noticeably, even if students work in pairs or small groups, the most common
48
49
50
51
52
53
54
55
56
57
58
59
60

Deleted: F
Deleted: of laboratory tasks have been analysed
Deleted: by Chinn and Malhotra (2002a) and t
Deleted: which has been
Deleted: e
Deleted: typically are given a
Deleted: problem to investigate
Deleted: and
Deleted: the
Deleted: a
Deleted: and
Deleted: some
Deleted: is a
Deleted: what

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

organisation for laboratory work, the overall picture is unchanged. Laboratory group work generally means students working alongside each other following teacher and task instructions, not conducting *group discussions*.

Despite this situation, scientific argumentation could be taught efficiently in the laboratory. Importantly, the argument is commonly turned the other way around, claiming that *if* students are doing investigations in the laboratory these *should* include argumentation; because, unless they do, the activities portray a misleading image of science (Driver et al., 2000). From this view, however, we deduce that training students to do investigations could be training in scientific argumentation. Hence, this rationale underpins intervention studies exploring ways of teaching scientific argumentation (Cavagnetto, 2010). Nevertheless, few studies utilise investigation contexts in which students collect their own data in a laboratory setting. Alternatives such as computer-based experiments (Clark & Sampson, 2008), observations from video (Engel & Conant, 2002) and analysis of a secondary data set (Sandoval & Reiser, 2004) are apparent. A common point, however, among all these studies *and* those using laboratories is that the investigation provides *scaffolding* for argumentation teaching. In other words, when teaching argumentation through scientific investigation teaching can be structured around debating alternative hypotheses, analysing data and evaluating evidence for a conclusion. This, we may claim, strengthens the case for teaching argumentation in the laboratory compared to many other contexts, because the scaffolding is already in place. As we have seen, problems occur when students mechanically follow ‘scripts’ in their laboratory investigations, but breaking out of this by ‘forcing’ students to reflect on and debate what they do and observe may provide a good way forward. Several intervention studies follow this line of argument. For example, Watson, Swain and McRobbie (2004), conducted an open-ended laboratory task running over three lessons in a UK school. They stimulated argumentation by guiding students towards ‘focusing’, ‘planning’, ‘obtaining evidence’, ‘interpreting’ and ‘evaluating’. In Israel, Katchevich, Mamlok-Naaman and Hofstein (2010) allowed students more time to plan open-ended investigations and encouraged discussion at various stages during their work. Both studies report positive outcomes compared to the ‘ordinary’ laboratory work Newton et al. (1999) describe. However, although these studies increased the amount of argumentation in laboratory investigations, both found the *quality* of argumentation to be relatively poor. Watson et al. (op cit) found students mostly met claims with counter-claims rather than using data from their investigation. Katchevich et al. (op cit) similarly report low level argumentation as measured by a coding framework. Two other US studies suggest improved argumentation quality depends on persistent use over a longer period (Richmond & Striley, 1996; Yerrick, 2000). Yerrick’s study was carried out over eighteen months. Pre- and post-intervention interview data reveals that the number of times students linked observational evidence with their proposed warrants more than doubled over this time period. Richmond and Striley made students do four investigations over a period of three months. All data gathering was qualitative, but in their conclusions they state:

One of the most significant changes we observed over the course of these four experiments was in students’ ability to formulate appropriate scientific arguments: They became more adept at identifying the relevant problem, collecting useful information,

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3 *stating a testable hypothesis, collecting and summarizing data, and discussing the*
4 *meaning of data.*

5 (Richmond and Striley, 1996: 847).

6
7 Besides offering structure for argumentation teaching, the laboratory benefits may also
8 accrue from the availability of students' data, which acts as a stimulus for debate. Any
9 teaching activity aiming to create debate depends on stimuli to generate different
10 opinions and contrasting assertions. Activities placed in a socio-scientific context, for
11 example, commonly utilise political and ethical controversies (Sadler, 2004). Students are
12 stimulated to engage in debate by taking and defending a stand. Cognitive conflicts or
13 misconception-based discussion can also act as stimuli (e.g. Naylor, Keogh, & Downing,
14 2007). In student-to-student discussions *one* authoritative voice is less likely and debate
15 is triggered because different conceptions are 'compete'. Data interpretation is a common
16 stimulus for argumentation (Sandoval & Millwood, 2007). Debate is created about
17 different hypotheses arising from the data and their value as evidence for stated
18 conclusions. This, however, they may work better when data is collected by the students
19 themselves rather than taken from other sources, because the ownership students get to
20 the data. No study is found to actually have tested the effect of this, but Simonneaux's
21 (2001) and Jiménez-Aleixandre et al. (2000) suggest that personal involvement with both
22 data and the arguments is important.

23
24 Thus, we see that scientific investigations conducted by students offer a *natural locus* in
25 science education for teaching argumentation. Firstly, the investigation manifests what
26 we mean by scientific argumentation; secondly, it provides a scaffold for teaching; and,
27 thirdly, it creates ownership to arguments by being based on students' own data.
28 Katchevich, Mamlok-Naaman, & Hofstein, (2010), Richmond & Striley (1996), Watson
29 et al. (2004) and Yerrick (2000) all attempt to improve argumentation in the science
30 laboratory by using strategies strengthening these aspects. Internationally, other
31 intervention studies, however, suggest that teaching should go further, and do more to
32 'stage' debate. In Korea, for example, Kim and Song (2006) stimulated argumentation in
33 an open-ended investigation by having students present reports for 'peer review', and
34 arranging discussions in which students acted as critics in a similar manner to scientists
35 attending a conference. In Norway, Kolstø and Mestad (2005) used peer reviewing of
36 laboratory reports as teaching approach. In the UK, Naylor, Keogh and Downing (2007)
37 used concept cartoons to create 'conflicts' before the investigation and a whole-class
38 session to debate findings afterwards. In the US, Martin and Hand (2009) combined
39 investigations with presenting and analysing claims in writing. These research studies are
40 not presented to permit precise comparisons of the exact amounts and/or quality of
41 argumentation. Nevertheless they offer the impression that these strategies are important
42 supplements to 'ordinary' student investigations. A pertinent question, however, is to
43 what extent such additional strategies can be sustained in laboratory teaching? Some
44 strategies span the laboratory activity over several lessons and restrict the number of
45 activities students can do within a school year. They also rely on science teachers feeling
46 comfortable about participation in this type of activity.

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

In the current study we isolate and explore effects of using various strategies when teaching argumentation in the school science laboratory. We built on traditions established in science education and stayed within the range of teaching approaches normally used by science teachers. The main focus was to explore issues of achieving better scaffolding of argumentation during an investigation and to probe how different forms of data stimulate debate.

Classroom cultures and the role of the teacher

Learning science is learning to talk science (Lemke, 1990). The specialist language of science not only includes declarative and procedural concepts, but also carries an ideological position with reference to certain epistemological criteria (Kelly, 2007). The process of becoming a scientist involves adapting to a scientific way of thinking and working by participating in science culture over time. Although few students become scientists, a similar process of enculturation is put forward as an ideal for any science classroom (Driver et al., 2000). Science teaching should offer a learning environment that promotes the values and norms of science and gives students rich opportunities to engage in scientific discourse. The typical discourse of the science classroom, however, is tightly controlled by the teacher, using what Lemke (op cit) refers to as the ‘triadic dialogue’; a question-answer-response pattern that gives little room for students to talk science and practice making the language of science their own. Many have therefore concluded that one of the biggest challenges of teaching scientific argumentation is to create a classroom environment less dominated by teachers holding on to a didactic teaching style (Herrenkol, Palincsar, DeWater, & Kawasaki, 1999; Jimenez-Aleixandre & Erduran, 2008; Kuhn, 1992; Zohar & Nemet, 2002).

Ideally, the teacher should fulfil two roles. One is *facilitator* encouraging students to participate in discussions by presenting views and critiquing those of others (Simon, Erduran & Osborne, 2006). When applied to investigative work, this means encouraging students to present and discuss alternative hypotheses and alternative ways of solving the tasks, rather than checking if they have the ‘correct’ solution. Students need to feel confident and able to accept that disputes are accepted and natural to scientific inquiry. The second role is to model good scientific practice. Jiménez-Aleixandre (2008) describes this as being an *able peer*, providing and persuading the use of scientific epistemic criteria. This can be done by engaging in debate and asking open questions aimed at eliciting justifications (Jimenez-Aleixandre, Lopez Rodriguez, & Erduran, 2005). The teacher may challenge students’ ideas, pointing out limitations and inconsistencies (Mork, 2005). These two roles, plead for an active teacher engaging as a ‘partner’ of student groups running their own debates.

The teacher’s importance, of course, may undermine the relevance of exploring strategies in investigative tasks. Firstly, if the classroom does not have an environment supporting argumentation, no teaching strategy is likely to prove efficient. Secondly, from a research methodological point of view, the classroom environment may be a confounding variable seriously distorting what we want to investigate. To compensate, the current study emphasised creating a learning environment in accordance with the ideals presented above and keeping this constant in all teaching.

The study

The study started by identifying common investigative tasks relevant to the England and Wales science National Curriculum (Qualifications and Curriculum Authority, 2007). Although variation exists, some tasks are used repeatedly in many classrooms, partly due to guidance from assessment authorities (Donnelly et al., 1996) and partly for practical reasons: finding good research or genuinely open-ended problems suitable for student investigations in school contexts is thought to be difficult. Good examples are easily established through science course materials, so tend to reoccur often in lessons. Typically, tasks require students to find the effects of one or more independent variables on a single dependent variable. Even if tasks are characterised as “open-ended”, students are often given the problem and follow standardised, “script”-based procedures to arrive at pre-determined solutions. Newton et al. (1999) observed such “open investigations” being used in four out of twenty-three randomly observed practical work lessons. Two typical tasks were chosen for this study (see below), selected partly because they are familiar to teachers but also because the equipment needed is very simple.

Deleted: in

Deleted: c

Deleted: teaching materials and

Deleted: As mentioned, typical

Deleted: are to

Deleted: a research

Deleted: 4

Deleted: 23

Deleted: of these

Investigations are at the heart of the problem of creating argumentation in the science laboratory: they are intended to introduce students to methods of scientific inquiry, but often fail to do so because of practical and pedagogical constraints (Chinn & Malhotra, 2002a). These activities involve no discussion about the evidence generated, or the way a conclusion has been drawn, since data are “unproblematic” and the answer “obvious”. Our next step was therefore to reflect on ways in which this pattern could be broken, changing tasks to prompt more authentic scientific argumentation. Working from the rationale suggested by the introduction and research studies promoting argumentation strategies in science lessons more generally (e.g. Osborne, Erduran and Simon, 2004), we identified the three strategies outlined below.

Deleted: The

Deleted: in the study

Deleted: and the

Deleted: changed

Deleted: include

Deleted: in

Deleted: looking towards

Deleted: possible

Deleted: e first strategy attempts to stimulate scientific argumentation by use of “complex data”; i.e. by

Deleted: .

Deleted: This strategy was built into a

Deleted: name for students

Deleted: . The task was set in the context of investigating transportation of hot liquid in metal tanks with different surfaces.

Deleted: with three different surfaces:

Deleted: coloured

Deleted: Being of the same material the only factor separating the containers was the surface colour

Deleted: this gives a s

Deleted: will have

Deleted: in temperature

Deleted: A

Deleted: leading

Deleted: to

Use of complex data

This involves leading students into a situation where data do not give one obvious “correct” conclusion. We adapted a task investigating the temperature drop occurring over time when containers of different materials are filled with hot water (Appendix 1, Task 1, “Container”) for this. The activity was set in the context of investigating transportation of hot liquid in metal tankers with different surfaces. Students were provided with metal containers (empty, clean food tins) coloured black, white and (shiny) grey metal. The material was identical, so only the surface colour varied. Small differences in temperature drop occur due to heat radiation. The shiny container exhibits least radiation and consequently the smallest temperature drop. As all temperature drops over time are small, students produce measurements in which differences in data collected from the three containers are of similar magnitude as the level of uncertainty in each measurement.

We anticipated that students would be led into and take different stances in a debate about data as evidence for claiming which container best retained the water temperature.

Conflicting hypotheses

Peer argumentation in the school science laboratory – exploring effects of task features

Here we presented students with conflicting theoretical hypotheses, about which they were likely to disagree. The students debated which of a set of statements were right or wrong before starting the investigation, and then to investigate the truth of these by experiment. Dissolving salt (sodium chloride) in water of different temperatures (Appendix 2, Task 2, Dissolving Salt) was the task. Initial statements were based on the common misconception among 11 – 16 year olds that sugar and salt lose mass when dissolved in water (Ebenezer & Erickson, 1996).

We anticipated that students would support different hypotheses, or, as a minimum, perceive the investigation as testing competing hypotheses, and engage in debate about these during phases of the investigation.

Post-investigation discussion

The third alternative differs from the first two in that a discussion is arranged after the practical investigation is complete. Handling equipment and gathering data are complicated tasks, so arranging a post-experiment discussion is an alternative to 'scaffolding' that may prompt students' reflecting about scientific inquiry. A task was developed that followed the "Container" (Task 1) practical investigation described above. Hence, this third task shared with Task 1 "Container" the element of using data to stimulate argumentation. Students were presented with fictional data related to the "Container" investigation collected by imaginary student groups who had used different methods (Appendix 3, Task 3 Post discussion). Questions directed students to evaluate the methods the data produced, and to debate evidence for the conclusion. Students answered questions individually and then in group debated in order to reach a common conclusion.

We anticipated that students would take a more holistic perspective on the investigation, looking at the methods for data gathering and the results when debating the final conclusion to be drawn.

The study investigated the effects of each alternative task on 12 - 14 year old students' argumentation when working in small groups. The research questions probed were:-

1. To what extent do these alternatives stimulate scientific argumentation?
2. If argumentation occurs, what forms of debate arise?

Design and methodology

The design was based on testing the three different strategies in ordinary whole class laboratory teaching. The tasks were planned for lessons lasting 60 minutes (a time which is common in secondary schools in England and Wales), at an independent school (i.e. non-state funded) in the North East of England. The participants were three Year 8 (12-14 year old) mixed ability classes, each with 22-23 students. The science teacher (author 2 of the paper) taught all classes. This ensured consistency across the teaching situations and a classroom climate that encouraged argumentation as indicated above.

Deleted: The second strategy

Deleted: s

Deleted: a

Deleted: are asked to

Deleted: a

Deleted: in

Deleted: s

Deleted: The strategy was implemented into another investigative task common in science lessons:

Deleted: d

Deleted: S

Deleted: the

Deleted: commonly found

Deleted: school students

Deleted: or

Deleted: s

Deleted: strategy

Deleted: first

Deleted:

Deleted: has been finished

Deleted: afterwards

Deleted: as

Deleted: help

Deleted: be more

Deleted: ve

Deleted: up

Deleted: in Task

Deleted: in the task were

Deleted: at making

Deleted: and

Deleted: started by

Deleted: ing

Deleted: ing

Deleted: their answers

Deleted: try

Deleted: strategy

Deleted: investigated

Deleted: strategies

Deleted: for each of these strategies

Peer argumentation in the school science laboratory – exploring effects of task features

Two classes carried out all three tasks in small groups. From these classes, four groups, each of three students, were selected for in-depth analysis. From the third class two groups were selected. These acted as “controls” by doing only Task 3, the “post-investigation” that is without first having conducted the matching practical task (1). Groups were sampled to mix gender and ability. Table 1 shows the data gathering design. The design made it possible to compare different student groups doing the same activity to see how a strategy worked for four different student groups. We were also able to compare the same group doing three different tasks to see how a student group reacted to three different strategies. The two groups doing only Task 3 enabled comparison with students who did this task with and without carrying out the practical task (Task 1) first. Data gathering was carried out over a period of two weeks.

(Table 1 in about here)

Data gathering was made with video recording, using one high quality Sony DVCAM camera on each focus group, and wireless “fly” microphones attached to the students.

This ~~enabled detailed recordings of student activities, and their conversation in a classroom with background noise. The microphones also recorded conversations between focus group students and the teacher. Using video camera and microphones, of course, will influence students’ talk, but students in all groups kept talking freely and their behaviour suggested no obvious constraints.~~

Students produced written reports, one ~~per group, from each task.~~ These were collected and used as background information to support and validate ~~video recording data.~~ In Task 3 students also produced written responses ~~for comparison.~~

The observed lessons ~~began with 5 – 10 minutes teacher introduction to the task.~~ Emphasis was placed on students being explicit about their thinking and ~~needing to make decisions during the investigations. They were also told that conclusions should be evidence-based and that achieving this should be a particular focus.~~ The students were used to working in the laboratory and familiar with investigative tasks. No specific teaching, however, had been provided on argumentation. The introduction and clearing up afterwards left 30 to 40 minutes for each practical investigation (Task 1 and 2). Task 3 required less time and lasted 10 to 20 minutes, depending on how quickly students handled the task.

Data analysis

Data analysis started with first transcribing all conversations from the video recordings. The transcripts were checked by a second, qualified academic colleague to ensure their correctness. The remaining data analysis was made from the transcripts, but often with a return to the video recordings to have a better understanding of the context or by confronting students’ written reports. Although this is a small scale study, the analysis was both qualitative and quantitative.

Identifying argumentation discourse

- Deleted:** made it possible to
- Deleted:** in detail what
- Deleted:** s
- Deleted:** did
- Deleted:** between the students
- Deleted:** spite of being in
- Deleted:** due to this situation
- Deleted:** each
- Deleted:** the
- Deleted:** investigations
- Deleted:** from the video recordings
- Deleted:** that were compared with the videos.
- Deleted:** s all started
- Deleted:** by the teacher
- Deleted:** the
- Deleted:** needed support from
- Deleted:** the

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The main focus of the analysis was to identify the type of argumentation discourse occurring in the student groups, but for comparison between tasks we also “measured” frequency and quality of argumentation. Several possible analysis methods were considered, as there is no one obvious method (Erduran, 2008). Watson (2004) counted claims made by students and looked at how often these were supported by data. Newton et al. (1999) focused more generally on the orientation of students’ work, e.g. “group discussion” and “closed experiments”, and measured time devoted to each of these. Students’ discussion in the different orientations were then studied qualitatively, and together with the quantitative time measure gave an indication of the amount of argumentation happening. A third approach, adopted here, is to identify “argumentation units” using Toulmin’s (1958) Argumentation Pattern (TAP). TAP describes argumentation with a claim being supported by data, and with various ways of strengthening or undermining this relationship. Each unit, of course, can last for various lengths of time, but this is subordinate to the quality of the argumentation contained. The argumentation measure is therefore given by number of units (more units indicate more frequent argumentative discourse) and the quality of each. Counting and scoring the units is not straightforward, because of the difficulty related to deciding what is claim, data and warrant (Erduran, 2008). When a new claim is made, for example, it may sometimes be judge as a simple “counter claim” (meeting a claim with another claim), and therefore belong to the same unit, or be seen as the start of a new unit. A depending factor is how it is supported with data and warrants. The numbers produced for units should therefore be read with some caution. They give meaning mainly when compared within the study: here, the researchers have been able to compare units and agree on definitions used consistently in the three tasks. In accordance with Zohar and Nemet (2002), claims made without any justification or not met with a counter claim were not recognised as an argumentation unit.

Determining the quality of argumentation discourse

The criteria used for scoring quality of argumentation units relates to low quality argumentation being “sparse” with few backing or rebutting elements and high quality argument “rich” in such elements. To score this issue we used a classification system developed by Erduran, Simon and Osborne (2004, see Table 2). Other frameworks were considered (an overview is presented in Sampson & Clark, 2008), but Erduran et al.’s (op cit) framework was selected because the five level scale provides a means of grading students’ comments. The key discerning factor is the presence or absence of rebuttals. In the two lowest levels there is no questioning of claims. Students either meet a claim with another claim (Level 1) or use some form of argument to support their claim (Level 2). The three next levels considered the quality and quantity of rebuttals. Identifying and coding argumentation units were equally important for the qualitative analysis of the group discussions as it was for the “scoring” and “counting”. The close inspection of units served as a means for understanding the nature of the laboratory discourse.

(Table 2 in about here)

Peer argumentation in the school science laboratory – exploring effects of task features

Two researchers identified units and carried out the coding for each piece of transcript. When units were agreed, inter-coder reliability was 70 - 80 % when coding their levels. Disparities were resolved through discussions.

Student orientation

An additional coding of students' "orientation", or focus, while working on the tasks was also conducted. This coding was not planned in advance but derived from the data as an attempt to characterise ways in which students solved the investigation tasks. Theoretically the coding has support in Klahr, Fay and Dunbar's (1993) *Scientific Discovery as Dual Search* (SDDS) model. This model suggests that someone (students and scientists) working on an inquiry task may operate in different "problem spaces". The model has two problem spaces, the *experimentation space* and the *hypothesis space*, but Klahr et al. add a third dimension, making in total three orientations:

- a) Experimentation: Students are focused on data gathering and handling of equipment
- b) Hypothesising: Students are focused on explaining the observed phenomena by use of scientific theories and concepts
- c) Co-ordination and evaluation: Students are focused on co-ordination and evaluation of evidence to draw a conclusion.

By using the SDDS model as a theoretical underpinning is suggested that students may be operating in these orientations disjointedly and with some problems combining them or moving from one orientation to another. Further details about how the coding was conducted will be given in the result section below.

Results

We will start by presenting two examples demonstrating students' complete work on Task 1 and 2. The examples illustrate characteristics of group work being "rich" (No 1) and "poor" (No.2) on argumentation. As such, these represent extremes in the findings. Number 1 is Group A carrying out Task 1 (Container, complex data) and Number 2 is Group C carrying out task 2 (Dissolving Salt, conflicting hypotheses). Students in the groups have been given fictitious names with first letters matching the group letter. Student comments are reported verbatim with editorial additions shown in square brackets to ensure the context is clear.

Example 1: Group A on Task 1, "Container" with complex data

After having been informed about the task and organised as a group the three girls in Group A went directly on to collecting equipment and setting up the experiment. No discussion occurred about the purpose of the task. They exchanged single comments as they went along with the preparations, repeating information from the task sheet or given by the teacher, to clarify and agree what they were doing. For example (brackets show time in minutes: seconds):

Ann (0:25): We need to explain clearly what goes into the container.

Peer argumentation in the school science laboratory – exploring effects of task features

Nearly nine minutes later, three metal containers with different surfaces (white, black and metal), each with a thermometer were lined up at the table and filled with hot water from a kettle. Then came explicit comments about the research question:

Ali (8:50): We need to do something..

Ann (9:03): We need to see which one ... like keeps the most heat..

It was clear that the students' focus so far had been *what* to do rather than *why*. Having made their first recordings of temperature in each container the discussion continued:

Ann (9:22): Do they [the company mentioned in the task] want to keep it hot .. or do they want to keep it cold?

Ada (9:29): I don't really know

Ali (9:29): They want to keep it warm

Ann (9:30): They want to keep it hot, okay

Ann (9:31): Okay, so we see which stays the hottest

Ali (9:32): This is the hottest at the moment...

This was their only explicit "planning" and in the following twelve minutes the students were fully engaged with recording data. As they went along, however, three more methodological issues (italicised) were brought up, all in the same accidental way and each was solved in a single sentence:

First, *how to present data*:

Ali (11:31): So.. we like, make a table

Ada (11:31): [Grunts] Yeah

Second, *what was the temperature of the water when they started?* (They forgot to record temperatures immediately after having poured the water.)

Ada (13:15): What was the temperature at zero Ali?

Ali (13:17): Temperature at zero was 100 degrees for everything, because it had all come out of the ..eh.. kettle.

Third, *how long should they keep taking measurements?*

Ali (21:15): How many minutes do we have to do it for?

Ann (21:17): I don't know

Ali (21:20): Just keep going until she (the teacher) stops us

Ann (21:24): Yeah

During the data gathering analysis was also carried out in a similar way, as the girls started to discuss which container best keeps the temperature. This started with Ali

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3 making a comment after the ninth set of readings, in which the temperature in the metal
4 coloured container (“silver”) was 86 °C, the black was 84 °C and the white was 81 °C:

5 Ali (18:47): I think silver is going to be the best, don’t you

6 Ada (18:49): Yeah

7
8 A few minutes later Ali declared the “winning container” and told the teacher they had
9 finished the investigation. The teacher wanted the students to evaluate the data before
10 drawing a conclusion, but the students did not see a need for this. When the teacher asked
11 the students *why* the metal surfaced container had best kept the temperature, it became
12 clear that the students had expected the black container “to win”:
13

14 Ada (24:58): [Be]cause it absorbs heat, like when the Sun is...

15 Ali (24:58): Like when you wear black clothes in the Sun

16 Ada (24:58): Yeah, in the Sun and then it absorbs the heat

17
18 These ideas, however, were never discussed among the students while doing the
19 investigation, or in the students’ report, which just presents the measurements and
20 confirms that the metal coloured container came out with the smallest fall in temperature.
21

Example 2: Group C, Task 2: “Dissolving” with conflicting hypotheses

22
23 The students in Group C, two boys and a girl, started by reading out the three statements
24 in the task (see Appendix 2) and discussing which were “right” or “wrong”. Although
25 lasting less than two minutes the discussion made a basis for the further investigation.
26 Two students (the boys Cameron and Callum) supported the statement “mass cannot
27 disappear” and claimed the opposite statement “there is always a small loss of mass”
28 therefore had to be wrong. The third student, Cynthia, was less sure and thought
29 statement B might be right. Statement C (mass depends on the temperature of the water)
30 was declared wrong by Callum. The two other students expressed some doubts about this,
31 but Callum’s self-assuredness silenced their views. The final outcome, decided mainly by
32 Callum, was to do experiments to test statements A and B.
33

34
35 The students decided on using 25ml water and 1 g salt to test hypothesis A. These were
36 measured independently before the salt was dissolved in the water and a new
37 measurement made of the solution. With small amounts, the measurements were not
38 exactly the same before and after dissolving the salt. There seemed to have been a loss of
39 mass, which confirmed Cynthia’s view:
40

41 Cameron (16:15): Some is gone... B is right, I think

42 Cynthia (16:16): I said B. Didn’t I say B!

43
44 At first they were all convinced about this conclusion, even Callum:
45

46 Callum (16:32): So there is always a small loss of mass when it dissolves...
47 The mass will increase, but less than the mass of the salt
48 added.
49
50
51

Peer argumentation in the school science laboratory – exploring effects of task features

1
2 Cameron (16:47): That is right.
3 Callum 16:47): So B is right.
4

5 Cynthia then suggested they had *disproved* statement A only and that they still had to
6 *prove* statement B. A 15 minute discussion started in which Callum tried to convince the
7 others that having a test disproving A is the same as proving B to be right. The outcome
8 of the discussion was an agreement that they should do the same test again. When doing
9 this they tried to copy the exact amounts used in the first experiment, arguing that “things
10 had to be kept the same”. This, however, was difficult and their next set of readings was
11 different:
12

13 Callum (21:26): Shouldn't it be the same as that [compares measurements
14 from the two experiments]
15 Cynthia (21:30): No, it wasn't, like, precise. You can't get ... or you can, but
16 it would be really hard to get exactly the same amount
17

18 This comment from Cynthia gave Callum a sudden understanding that hypothesis A still
19 might be right and that the unexpected results in the first experiment was due to
20 measurement error:
21

22 Callum (21:36): No, wait a moment (eager). That means that that [statement
23 A] could be right. Because, it is really hard to get the exact
24 measurements.
25

26 He continued:
27

28 Callum (21:56): I do not think we lose anything. It is just impossible to
29 make it precise.
30

31 Callum tried to convince the others about this. They, however, still thought they should
32 stay with the outcome of the measurements, which indicated a loss of mass (i.e.
33 supported statement B). The discussion included detailed analysis of the experiment and
34 why or why not the final outcome should be trusted. Callum did not manage to convince
35 the others on his view.
36

Patterns of orientation in students' investigations

37
38 The two examples show different approaches to the investigations. Example 1
39 demonstrates an algorithmic approach, focusing on data gathering to answer a question.
40 The girls implicitly “knew” what to do, not needing planning, and they unreservedly
41 trusted the outcome of the measurement, not perceiving a need to evaluate. A conclusion
42 could be drawn, and the investigation was “finished” as soon as the last measurement was
43 made. Example 2 developed differently. More time was spent on planning and on
44 conceptual discussion, but the most characteristic difference from Example 1 is extensive
45 discussion about what conclusion could be drawn from the data. This was absent in the
46 first group, but filled more than half the time of the second group. The difference has an
47 obvious importance for occurrence of scientific argumentation.
48
49
50

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

As a way of demonstrating these differences across all groups and tasks a decision was made to code time spent on the different orientations of students' discourse. Social talk, talk about the task (rather than problems in the task) and discussing/writing the report were excluded from this analysis and coded as "other", as these were not our focus. The transcript was the main source of coding but with support from the video pictures. Sequences of the group investigation were given a code a) Experimentation, b) Hypothesising or c) Co-ordination and evaluation, and the time spent on each of these was added up.

Figure 1 shows how time was spent by the groups on each of the orientations over the tasks they carried out, as shown in Table 1. The category "other" is excluded and the percentage is therefore a relative distribution of time spent between the three main categories.

[INSERT Figure 1 about here]

The graphs show that Task 1, Container with complex data, prompted students to focus on "experimentation". Groups A (the first example presented above), held this focus 90% of the time, hardly devoting any time to "hypothesising" and "coordination and evaluation". Group C on Task 2, Dissolving Salt with conflicting hypotheses, (the second example presented above) is an exception, spending more time on coordination and evaluation than on experimentation. Task 2 stimulated more discussion about co-ordination and evaluation than Task 1. However, we sense some "randomness" in the data: how a task develops for the groups is inconsistent. Groups B and D exhibit similar patterns for Tasks 1 and 2, indicating that orientation is group-related (groups having the same orientation across tasks), while the opposite is the case for Groups A and C, indicating that orientation is task-related. All in all, this indicates a situated effect, suggesting that the way in which a task develops for a pupil group is influenced by factors happening during that particular event.

Task 3, the post-investigation discussion, naturally has a higher percentage of the time on the two last categories, since it was non-practical. We, see that students, except for Group D, balance their time between "hypothesising" and "coordination and evaluation". This indicates that hypothesising, which was obviously very difficult to stimulate during the practical work, was more easily stimulated in this task format.

Argumentation and students' orientation

Table 3 summarises the number of argumentation units in each of the three tasks. We see that Task 2, Container, had more units than the two other tasks, generated by groups A and C. Task 3, the post-investigation discussion of Container, had fewest units. It is, however, important to keep in mind that Task 2 lasted 30-40 minutes while Task 3 was solved within 10 to 20 minutes

(Table 3 about here)

Peer argumentation in the school science laboratory – exploring effects of task features

Figure 2 presents the level of the argumentation units in Table 3 when scored against Erduran et al.'s (2004) framework. We judged most argumentation at level 2, in which students present a claim with some form of justification, but without rebuttals. Tasks 2 had more units at higher levels than the other tasks, stimulate by 'conflicts' established in the pre-discussion. In discussions trying to solve these conflicts later in the investigation (as demonstrated in Example 2 above) students made use of data and presented rebuttals. This brought the argumentation to higher levels.

(Figure 2 in about here)

Deleted: Tasks 2 and 3 have relatively more units at higher levels than Task 1.

Figure 3 presents the number of argumentation units in each of the three "orientations" presented earlier, but summed up across all tasks. 12 % only (10 out of 81) happened in "experimentation", which took on average more than eighty percent of students' time in the practical tasks. Argumentation occurred most frequently when students were coordinating and evaluating the evidence and conclusions.

(Figure 3 in about here)

Figure 4 suggests that argumentation unit quality was enhanced to 2.7 on average when students were "hypothesising" and doing "coordination and evaluation", compared to 1.8 for "experimenting". Argumentation in the "experimentation" mode meant mainly students making claims in relation to the data recording. For example, in Task 1, claiming one container to be "best" and supporting this with the measurements of temperature. A claim like this was unlikely to be rebutted by another student and the argument unit was very short. Some discussion occurred about methods and triggered some more advanced structures, but this was rare and also relative short. For example, Darren in Group D made the claim that their experiment on dissolving salt (Task 2) was "false":

Darren (12:03): ..we have changed the amount of water since the first time we weighed it, so this mass will be totally different. So this experiment is false.

Some discussion developed from this, but the problem was solved by carrying out the experiment again. The argumentation occurring when "hypothesising" had higher levels because of students rebutting each others explanations, but this again was not a very frequent event. In the "coordination and evaluation" mode the type of argumentation had much more variation. Students could go into detailed discussion about data in order to solve disputes about what conclusions to draw. This has already been demonstrated by the examples given in the Group C solving Task 2, but a more detailed illustration of the type of argument will be demonstrated. This is Group D discussing the statement that "mass is dependent on temperature of the water" in the same task. Darren made the claim that temperature has no effect, but was challenged with the rebuttal that "salt was at first observed at the bottom of the beaker, but disappeared when the temperature increased"; based on the warrant that solid salt and dissolved salt do now weigh the same. Darren first agrees with the observation but attacks the warrant:

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Darren (24:00): The higher the temperature the more salt can dissolve, but it does not affect the mass; unless you put more salt in. If you have the same amount of salt and the same amount of water, heating doesn't actually change the actual mass that is in there

He then strengthen is own argument with a rebuttal

Darren (24:25): Unless the water is evaporating.

Although the values in Figure 2 to 4 should be read cautiously, considering the subjective character of identifying and scoring argumentation units, the results have a clear trend: the data gathering part of laboratory tasks does not invite argumentation discourse. Such discourse happens mainly when students try explaining observations and reflecting on the evidence for their conclusions.

(Figure 4 in about here)

What effects on argumentation relate to the strategies implemented in the tasks?*Complex data: Task 1, Container*

Considering that the temperature differences identified in the groups were 1 - 3 °C and that students sometimes disagreed when reading the thermometer, there were good grounds for discussion about the evidence for deciding which container had smallest heat loss. However, all groups concluded that "silver was best" regardless of how small the temperature differences were. When challenged by the teacher students did point towards "accuracy" of their data gathering. Their response reveals that "accuracy" was not interpreted in terms of *measurement error*, but rather as *measurement strategy*. They defended their conclusion with comments about measurement frequency and regularity, but never doubted that a single measurement could be wrong or uncertain. Hence, complex or "uncertain" data in itself did not create any discussion: there was no need to discuss evidence or the conclusion since measurements gave "the correct answer".

Conflicting hypotheses: Task 2, Dissolving Salt

This strategy prompted initial discussion as intended, bringing forward conflicting views and predictions. The students, however, soon put the conceptual discussion aside and focused on data gathering. Figure 1 shows a small proportion of time spent on "hypothesising" relative to "experimentation". Data gathering took most time and in this "mode" little attention was paid towards other parts of the investigation. Figure 1 also shows that more time in this task was spent on "co-ordination and evaluation" compared to Task 1, due to conflicts occurring when data did not give the predicted answer. Students supporting statement A (no loss of mass) expected measurements before and after dissolving the salt to be *exactly* the same, without considering differences due to

Peer argumentation in the school science laboratory – exploring effects of task features

measurement error, and this problem had to be solved. Several outcomes occurred, which aligns with the findings in Chinn and Malhotra (2002b):

- Group A denied the problem and simply accepted their measurements without saying whether they supported or refuted the hypothesis. That is, the data became the final answer to the task.
- Group B rejected their original hypothesis and concluded that salt loses mass when being dissolved in water, thus accepting the data as “true”;
- Group C at first rejected their original hypothesis, but then discussed the evidence;
- Group D ignored anomalous data and kept the original hypothesis, thus accepting the hypothesis as “true” over the data.

Each of these situations gave a different ground for argumentation. If, as in group B, students believed in the data, argumentation was short:

- Ben (24:15): So we were wrong.
Bob (24:29): 183.69 [reads out the mass of the dissolved salt in solution]
Ben (24:33): Did you say it should be exactly....?
Bea (24:40): We might have missed out a bit of salt
Bob (24:55): So we might lose a bit of mass, so the hypothesis might be right
Bob (24:58): Our hypothesis was incorrect!
Bea (25:05): So it will increase a little bit, but not as much as

Bea and Bob here raise a rebuttal, but these carry little weight because data are “correct” and give the “final answer”. Group D held similar certainties about the hypothesis, ignoring the data because they *knew* their hypothesis was correct. Only Group C, who accepted data as uncertain, argued with higher frequency and advanced into more use of rebuttals. Uncertainty in the data opens up the possibility that a hypothesis might be seen as correct or not, depending on data quality.

Two conclusions arise from the conflicting hypotheses task. Firstly, presenting alternative hypotheses stimulates argumentation. This happened especially in the initial phase when students tried to resolve conflicts between hypotheses and in the experimental phase when discussing matches between hypotheses and data. Such conflicts also occurred in Task 1, where students observed the metal-surfaced container best keeping the temperature despite believing the black container was “correct”, but this task did not initiate the same amount of argumentation. We attribute the difference to the explicitly-stated hypotheses in Task 2 that prompted initial discussion. Secondly, the frequency and quality of argumentation is strongly influenced by students’ understanding of uncertainty in the data. Students’ belief in the “truth” of data limits the quantity and quality of argumentation.

Post-investigation discussion: Task 3

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

This format prompted a focused and effective discussion generating more argumentation units per unit time about the “container” investigation data. The outside laboratory setting enabled students’ attention to be readily directed towards reflecting on both data and method. Surprisingly, students also paid more attention to “hypothesising”, although this was not explicitly required. Figure 1 shows almost all groups used 50% or more of the time explaining *why* a metal-surfaced container should keep temperature better than the white and black ones. This matter was neglected by students doing Task 1. The alternative hypotheses discussion stimulated argumentation with frequent use of warrants and rebuttals. Group E, for example, suggest first that the explanation lies in the property of the metal, but raised the rebuttal that all containers were the same type of metal. They pointed out that black is known to “attract” heat. Next they suggest “layers of the paint” as an explanation, but again found a rebuttal by consulting the data:

Eric (21:27): black might have had more paint on, but metal still does better.

Their last suggestion was that the temperature differences may be caused by “different physical properties of the paint”, to which they did not manage to form any rebuttal and therefore kept as their final explanation.

This task also showed that doing the related practical task (Task 1) first did not generate more or “better” argumentation (see Table 1). In fact, some group discussions indicate the opposite. Groups A to D, who did Task 1 first, compared the fictitious Task 3 data with their own investigation and results, so did not see a need for discussion: their own method, measuring temperature every minute, was judged the “best” strategy and their own data offered the “correct” conclusion. Groups E and F, who had not done Task 1 beforehand, discussed the four sets of data provided, drew a conclusion and used methodological criteria to decide which was experiment was best.

Deleted: These observations indicate that students collecting their own data may *restrict* rather than stimulate scientific argumentation. ¶

Discussion

Changing laboratory teaching from a “positivist” tradition (Driver et al., 2000), towards nurturing authentic scientific inquiry was the motivation for this study. To date, scientific argumentation, which should be a natural part of any inquiry process, has played a minor role in laboratory teaching. The oversimplified methods implemented in many laboratory tasks guide students from research problems to final conclusions without the need to raise questions about the method used or the quality of evidence collected. Consequently, data gathering becomes the main focus for student activity, while other elements of the inquiry process are neglected (Abrahams & Millar, 2008; Newton et al., 1999; Watson, 2004). Extant research has also shown that school experiments, for the above reason, teach students a misleading picture of scientific inquiry, reinforcing an unscientific epistemology by encouraging the belief that science is a simple, algorithmic form of reasoning (Chinn & Malhotra, 2002a; Germann, Haskins, & Auls, 1996). Our study challenges this practice by testing the effects of three different strategies in investigative tasks.

The findings point to difficulties in changing existing laboratory teaching practices. Despite working in a learning environment that strongly encouraged evaluation and

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

discussion of the quality of the evidence, and the task strategies, more than eighty percent of students' time and attention were focused on data gathering. Our evidence shows that this "mode of working" is the least stimulating for scientific argumentation. The three strategies diverted students' attention towards "hypothesising" and "coordination and evaluation" with mixed success. Students ignored complex data (Strategy 1) in Task 1 (Container) identifying the "best" container, without thinking of the possibility that their data may not fit this purpose. Conflicting hypotheses (Task 2, Dissolving Salt) produced more argumentation than Task 1, but did not for all groups. Only one group examined the evidence thoroughly, by accident, because they measured twice with different outcomes. Post-investigation discussion (Task 3) generated sufficient distance from the investigation to enable students to consider evidence in relation to the conclusion. But, even then, some students solved the task as looking for the "correct" approach and the "right" conclusion. Jiménez-Aleixandre et al. (2000) suggest that hypothetical data provided by the teacher generate different patterns of discussion than empirical, uncertain data, gathered by students themselves. A conclusion from this study is that students' working from their own data does not *guarantee* their engaging in evaluative debate.

Formatted: Font: Italic

The study points towards two key issues in understanding why it is so difficult to engage students in scientific argumentation while working in the laboratory. First, it support Kelly's (2005) claim that science inquiry is an epistemic practice which requires understanding of the methodological components and criteria involved. Our findings reveal how students take data to be "true", with no concept of "uncertainty". We were surprised by the extent to which students put aside personal beliefs to accept a measurement uncritically. Even when admitting an error, the conditions under which the measurement was made rather than the measurement itself were responsible - a measurement was always right. Watson (2004:34) similarly observed:

The students carried out their tests and accepted the results of their tests as proof: the results justified claims in an unproblematic way.

This has tremendous implications for any debate regarding evaluation and coordination of evidence: if data are "true", there is no need for argumentation. Interestingly, at times students could imply they understood that data had uncertainty, but their interpretation of "uncertainty" did not focus on measurement error, but on procedures for data gathering, such as measuring at regular intervals. On one hand, this demonstrates a failure of laboratory teaching: students learn procedures mechanistically, using them without epistemological understanding. Early laboratory teaching should focus on epistemological understanding and less on experimental procedures. On the other hand, it also demonstrates that the strategies trialled in this project were not sufficient to change students' understanding of the epistemic nature of scientific inquiry.

Second, of significant importance for laboratory-based scientific argumentation, is the restraint relating to "working modes". Klahr et al.'s (1993) SDDS model offers theoretical support for this. Their "problem space" concept suggests scientists operate in three separate modes, "experimenting", "hypothesising" and "coordination and evaluation", where the last mode requires a combination of the two first. This model had

Peer argumentation in the school science laboratory – exploring effects of task features

1
2 strong support in our data. Empirically, however, it was surprising to find how strictly
3 students kept to one of the modes and how difficult they found it to change from one to
4 another. Students spent most time in “experimentation” mode, which may seem natural as
5 they are working with equipment and asked to do data gathering. “Hypothesising” could
6 have been a natural mode, but was rarely found while students were working with the
7 equipment. Only when they put the equipment aside and sat down to discuss the
8 investigation, as in Task 3 was this mode apparent. The reason for this, we believe, is that
9 mental effort is required to leave one mode for another. Operating in a mode means
10 seeing the task in a particular way, being aware of possibilities and restrictions for what is
11 relevant and important. To enter another mode involves changing the conception of the
12 task, so needs some stimuli. This applies when going from “experimentation” to
13 “hypothesising”, but is even more relevant for “coordination and evaluation”. While the
14 two first modes are familiar to students, this last mode is relatively unknown. Many
15 students looked rather puzzled when being asked to evaluate their conclusions and lacked
16 a strategy for this. Familiarising students with the mode of coordination and evaluation
17 may be a start towards establishing argumentation in the laboratory. That means
18 establishing some understanding of the *purpose* of the task and demonstrating some
19 examples for strategies to handle the problems involved.
20

21 It should be noted, of course, that the problem of engaging students in argumentation
22 while being in the laboratory could be solved partly by creating more debate in the
23 experimentation mode. Students could be made to reflect on and debate if the methods
24 they use and the data they gather are accurate and give the right information. However,
25 for two different reasons this can only be part of the solution. Firstly, students’
26 investigations have to be kept simple. If the design of an investigation gets too
27 complicated students lose track and their work becomes meaningless. We observed
28 examples of this during the study. Data have meaning only if they are made with full
29 understanding of the design and method from which they originate. Simple investigation
30 designs also limit how much debate could be prompted. Secondly, the argumentation we
31 aim for goes beyond ‘experimentation’. Each problem space has its own rationale for
32 argumentation. When ‘experimenting’, the problem is how to make relevant and accurate
33 observations and collect data to answer the research question. When ‘hypothesising’, the
34 problem is to find the best explanation. When doing ‘co-ordination and evaluation’ the
35 problem is to decide the certainty of the explanation/conclusion. Doing an investigation
36 answers all these three problems. If we were training laboratory technicians we could
37 claim the first to be most important, but as students are being trained to understand
38 disciplinary critique (Ford, 2008) the two last problems that should take priority.
39

40
41 Applying these perspectives to the tasks used here, and in combination with the data
42 analysis outcomes, we see potential for further development. Generally, laboratory
43 teaching should work with, not against, the problem spaces. This means we should guide
44 and stimulate students’ discussion and debate towards particular questions at points in the
45 investigation where these might naturally occur. Task 1 (Container), for example, could
46 have had a pre-discussion similar to the ‘post-discussion’ applied in Task 3. That is,
47 students could have had several data sets as stimuli for discussing the best way of solving
48 the task and a likely conclusion *before* planning and conducting their own investigation.
49

This might open up rather than restrict discussion, as the analysis of Task 3 reports. This also ensures students approach data gathering with expectations, as Task 2 (Dissolving Salt) stimulated debate. Questions asked of students while data gathering should focus on technical issues only. A post-discussion added to Task 1, with an explicit focus on the difference between explaining and evaluating the data would also be valuable. Hence, we can combine the best elements of all three tasks used in the study, since none alone gives the best means of stimulating argumentation in the laboratory.

Although argumentation in a laboratory context has unique features, similarity is found with research conducted in other contexts. Roth and Roychoudhury (1992), De Vries, Lund and Baker (2002), Sherman and Klein (1995) and Zohar and Nemet (2002) suggest that a structured approach is important in generating productive discussion. Students need help when putting forward and identifying claims and evaluating using scientific knowledge and data. Other research suggests that establishing argumentation structures in teaching takes time, and over two weeks only limited improvement may be expected. Osborne, Erduran and Simon (2004) report their nine month intervention to be too short for students to develop sufficient skills and abilities. This is supported by Zoller, Ben-Chaim, Pentimalli and Borsese (2000), investigating first year college undergraduates development of critical thinking. They point towards the need of recurrent opportunities to engage in same type of activities. Only when students have obtained some understanding of the pattern and purpose of activities may we expect the teaching to become efficient. The message, however, from these and other studies is that teaching of argumentation *is possible if it is explicitly addressed and taught*. The current study indicates that the same applies to the science laboratory, offering a sought for opportunity to practice argumentation in a scientific, as opposed to the more common socio-scientific, context.

The difficulty of establishing argumentation in the laboratory, of course, should not be underestimated. Laboratory teaching is deeply rooted in the logical-empiricist tradition and has a long way to go before socio-constructivist epistemological perspectives become commonplace. If, as the current research suggests, this is hindered by the psychological constraints of students operating in different “working modes”, we should not be too optimistic for rapid change. More research is needed to establish better understandings of efficient strategies. The present study emphasised “implicit” teaching of epistemological criteria, and a useful alternative to be explored is what happens if these are made more explicit. A longitudinal intervention may provide revealing information. Laboratory work needs to prove less situated with “random incidents” being the main pattern. The hope is that recurrent use of task strategies may establish structures and understanding that students carry with them from one activity to the next. A last suggestion is the need for wider use of task strategies. Knowing that a teacher’s ability to foster a context of argumentation is crucial (Osborne et al., 2004), the current research, with teaching being conducted by one of the researchers, has clear limitations. The strategies should therefore be tested with other teachers and across different educational cultures. This and other research, however, should be given priority as the laboratory is too important as an arena for training scientific argumentation to be left unattended.

References

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Abrahams, I., & Millar, R. (2008). Does Practical Work Really Work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1464-5289.
- Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honour of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum.
- Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences*, 2(2), 141-178.
- Cavagnetto, A. R. (2010). Argument to Foster Scientific Literacy: A Review of Argument Interventions in K-12 Science Contexts. *Review of Educational Research*, 80(3), 336-371.
- Chinn, C. A., & Malhotra, B. A. (2002a). Epistemologically Authentic Inquiry in Schools: A theoretical Framework for Evaluation Inquiry Tasks. *Science Education*, 86, 175-218.
- Chinn, C. A., & Malhotra, B. A. (2002b). Children's Responses to Anomalous Scientific Data: How Is Conceptual Change Impeded? *Journal of Educational Psychology*, 94(2), 327-343.
- Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, 45, 293-321.
- De Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences*, 11, 63-103.
- Donnelley, J., Buchan, A., Jenkins, E., Laws, P., & Welford, G. (1996). *Investigation by Order. Policy, curriculum and science teachers' work under the Education Reform Act*. Nafferton: Studies in Education.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-312.
- Ebenezer, J., & Erickson, G. (1996). Chemistry Students' conceptions of Solubility: A Phenomenography. *Science Education*, 80(2), 181-201.
- Engel, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*(20), 399-483.
- Erduran, S. (2008). Methodological Foundations in the Study of Argumentation in Science Classrooms. In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education. Perspectives from Classroom-Based Research*. Dordrecht: Springer
- Erduran, S., Simon, S., & Osborne, J. (2004). Tapping into Argumentation: Developments in the Application of Toulmin's Argument Pattern for Studying Science Discourse. *Science Education*, 88, 915-933.
- Ford, M. (2008). Disciplinary Authority and Accountability in Scientific Practice and Learning. *Science Education*, 92, 404 – 423.

Deleted: .

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. *Journal of Research in Science Teaching*, 33(5), 475-499.
- Harmon, M., Smith, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V. S., Gonzalez, E. J., & Orpwood, G. (1997). Performance assessment in IEA's third international mathematics and science study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation and Educational Policy. Boston College.
- Hegarty-Hazel, E. (Ed.). (1990). *The Student Laboratory and the Science Curriculum*. London: Routledge.
- Herrenkol, L. R., Palincsar, A. S., DeWater, L. S., & Kawasaki, K. (1999). Developing scientific communities in classrooms: A sociocognitive approach. *The Journal of the Learning Sciences*, 8(3 & 4), 451-493.
- Hodson, D. (1993). Re-thinking Old Ways: Towards A More Critical Approach To Practical Work In School Science. *Studies in Science Education*, 22, 85-142.
- Hodson, D. (1996). Practical work in school science: exploring some directions for change. *International Journal of Science Education*, 18(7), 755 - 760.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: foundation for the 21st century. *Science Education*, 88, 28-54.
- Hofstein, A., Kipnis, M., & Kind, P. M. (2008). Learning in and from science laboratories: enhancing students meta-cognition and argumentation skills. In C. L. Petroselli (Ed.), *Science Education Issues and Developments* (pp. 59-94). New York: Nova Science Pub.
- Hogan, K., & Maglienti, M. (2001). Comparing epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching*, 38, 663-687.
- Jimenez-Aleixandre, M. (2008). Designing Argumentation Learning Environments. In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education Perspectives from Classroom-based Research*. Dordrecht: Springer.
- Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or "Doing Science": Argument in High School Genetics. *Science Education*, 84, 757-792.
- Jimenez-Aleixandre, M., & Erduran, S. (2008). Argumentation in Science Education: An Overview. In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education. Perspectives from Classroom-based Research*. Dordrecht: Springer.
- Jimenez-Aleixandre, M., Lopez Rodriguez, R., & Erduran, S. (2005). Argumentative quality and intellectual ecology: A case study in primary school. National Association for Research in Science Teaching, Annual Meeting. Dallas, TX.
- Katchevich, D., Mamlok-Naaman, R., & Hofstein, A. (2010). Argumentation in the Chemistry Laboratory: Inquiry and Confirmatory Experiments, NARST Annual Conference. PA: Philadelphia.
- Kelly, D. L. (2007). Discourse in Science Classrooms. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education*. Mahwah, NJ: Lawrence Erlbaum.

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3 Kelly, G. (2005). Inquiry, Activity, and Epistemic Practice, NSF Inquiry Conference
4 Proceedings <http://www.ruf.rice.edu/~rgrandy/NSFConSched.html> . , Inquiry
5 Conference on Developing a Consensus Research Agenda Rutgers University.
6
7 Kim, H., & Song, J. (2006). The Features of Peer Argumentation in Middle School
8 Students' Scientific Inquiry. *Research in Science Education*, 36(3), 211-233.
9
10 Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A
11 developmental study. *Cognitive Psychology*, 24(1), 111-146. *Cognitive*
12 *Psychology*, 24(1), 111-146.
13
14 Kolstø, S. D., & Mestad, I. (2005). Learning about the nature of scientific knowledge: the
15 imitating-science project. In K. Boersma, M. Goedhart, O. D. Jong & H. Eijkelhof
16 (Eds.). *Research and the Quality of Science Education*, 247-258. Dordrecht:
17 Springer.
18
19 Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*.
20 Cambridge, MA: MIT Press.
21
22 Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, 62, 155-178.
23
24 Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: Univ. of Chicago
25 Press.
26
27 Lemke, J. (1990). *Talking Science: Language, Learning, and Values*. Norwood, NJ:
28 Ablex.
29
30 Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Learning and Teaching in the
31 School Science Laboratory: An Analysis of Research, Theory, and Practice. In S.
32 K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education*.
33 Mahwah, NJ: Lawrence Erlbaum Associates.
34
35 Martin, A., & Hand, B. (2009). Factors affectin the implementation of argument in the
36 elementary science classroom. A longitudinal study. *Research in Science*
37 *Education*, 39, 17-38.
38
39 Millar, R., Lubben, F., Gott, R., & Duggan, S. (1994). Investigating in the school science
40 laboratory: conceptual and procedural knowledge and their influence on
41 performance. *Research Paper in Education*, 9(2), 207 — 248.
42
43 Mork, S. M. (2005). Argumentation in science lessons> Focusing on the teacher's role.
44 *Nordic Studies in Science Education*, 1(17-30).
45
46 Murphy, P., & Gott, R. (1984). *The Assessment Framework for Science at Age 13 and*
47 *15*. APU Science report for teachers: 2. UK: DES.
48
49 Naylor, S., Keogh, B., & Downing, B. (2007). Argumentation and Primary Science.
50 *Research in Science Education*, 37, 17-39.
51
52 Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the
53 pedagogy of school science. *International Journal of Science Education*, 21(5),
54 553-576.
55
56 Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in
57 school science. *Journal of Research in Science Teaching*, 41(10), 994-1020.
58
59 Qualifications and Curriculum Authority. (2007). *Science. Programme of study for key*
60 *stage 3 and attainment targets*. The National Curriculum 2007 (pp. 206-219).
London: QCA.
61
62 Richmond, G., & Striley, J. (1996). Making meaning in classroom. Social processes in
63 small group discourse and scientific knowledge building. *Journal of Research in*
64 *Science Teaching*(33), 839-858.

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Roth, W. -M., & Roychoudhury, A. (1992). The social construction of scientific concepts or the concept map as conscription device and tool for social thinking in high school science. *Science and Education*, 76, 531–557.
- Sadler, T. D. (2004). Informal Reasoning Regarding Socioscientific Issues: A Critical Review of Research. *Journal of Research in Science Teaching*, 41(5), 513-536.
- Sadler, T.D., Chambers, F.W. & Zeidler, D.L. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education*, 26(4), 387-409.
- Sampson, V., & Clark, D. B. (2008). Assessment of Ways Students Generate Arguments in Science Education: Current Perspectives and Recommendations for Future Directions. *Science Education*, 92(3), 447-472.
- Sandoval, W. A., & Millwood, K. A. (2007). What can argumentations tell us about epistemology? In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education*. Dordrecht: Springer.
- Sandoval, W. A., & Reiser, B. J. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3), 345-372.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. New York: Erlbaum.
- Séré, M.-G., Leach, J., Niedderer, H., Psillos, D., Thierghien, A., & Vicentini, M. (1998). *Improving Science Education: issues and research on innovative empirical and computer-based approaches to labwork in Europe*. Final report from Labwork in Science Education. .
- Sherman, G. P., & Klein, J. D. (1995). The effects of cued interaction and ability grouping during cooperative computer-based science instruction. *Educational Technology Research and Development*, 43, 5–24.
- Simon, S., Erduran, S., & Osborne, J. (2006). Learning to Teach Argumentation: Research and development in the science classroom. *International Journal of Science Education*, 28(2-3), 235-260.
- Simonneaux, L. (2001). Role-play or debate to promote students' argumentation and justification on an issue in animal transgenesis. *International Journal of Educational Research*, 23, 902-927.
- Tapper, J. (1999). Topics and manner of talk in undergraduate practical laboratories *International Journal of Science Education*, 21(4), 447 – 464
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Watson, J. R., Swain, J. R. L., & McRobbie, C. (2004). Students' discussion in practical scientific inquiries. *International Journal of Science Education*, 26(1), 24-45.
- Yerrick, R. K. (2000). Lower track science students' argumentation and open inquiry instruction. *Journal of Research in Science Teaching*, 37, 807-838.
- Zeidler, D., Sadler, T., Simmons, M., & Howes, E. (2005). Beyond STS: A research-based framework for socioscientific issues education. *Science Education*, 89, 357–377.

Peer argumentation in the school science laboratory – exploring effects of task features

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39, 35-62.

Zoller, U., Ben-Chaim, D., Pentimalli, R., & Borsese, A. (2000). The disposition towards critical thinking of high school and university students: An inter-intra Israel-Italian study. *International Journal of Educational Research*, 22, 571-582.

For Peer Review Only

Appendix 1 Task 1, Container

Container

A company transporting hot liquid is buying a new tanker lorry and wants to know what colour the tank should be. You are asked to do an experiment to help give them advice if the colour has any effect on loss of heat.

You should find out *if* the colour has any effect and, if so, *what colour* (black, white or metal) best keeps the temperature.



The company wants to be sure about its decision and therefore would like each scientist to tell what he or she thinks about the experiment. You should:

1. Do the experiment to get the data.
2. Discuss how big the differences are between the containers and how sure you can be that the results are correct.
3. Make an advice to the company, telling them if you all agreed or if there were different opinions.

Your answer to the company must be supported by data and you should try to explain the results.

You will have:

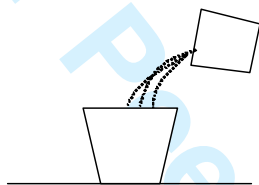
- Three types of containers
- Hot water
- Thermometers
- A timer

Appendix 2 Task 2, Dissolving Salt

Dissolving salt

Some students are investigating what happens to the mass of a container with water when we dissolve salt in it. They have different explanations. With whom do you agree? You should:

1. Discuss each statement and tell **if and why** you think it is right or wrong.
2. Select two statements you are uncertain about or you think are wrong.
3. Then do the experiments to test these two statements.
4. Discuss if the experiments are good enough to draw a conclusion and, if so, why or why not.



Statements:

- a) *When salt dissolves it still weighs the same. The total mass is therefore exactly the same as the water and the salt added.*
- b) *There is always as small loss of mass when something dissolves. The mass will increase but less than the mass of the salt added.*
- c) *The higher the temperature the more salt dissolves, the new mass therefore depends on the temperature of the water.*

You will have:

- Containers
- Hot and cold water
- Thermometers
- A stirrer
- A scale

Appendix 3 Task 3 Post-investigation discussion of Container task

Container

Some students have been doing an experiment to find out if the colour of an object has any effect and on its heat loss, and if so, what colour best keeps the temperature.

They had containers in three different colours: black, white and metal (not painted). In their experiments they filled the containers with hot water and measured the temperature as the containers cooled down. The tables show results from four different groups. They all did the experiment in different ways.

Group A			
150 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
1	90	90	91
2	84	84	85
3	80	79	80
4	76	75	76
5	72	72	73
6	71	70	72
7	70	69	71

Group B			
100 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
5	75	76	78
10	55	57	48
15	40	42	44
20	30	34	36
25	30	27	30
30	23	24	26
35	19	18	19

Group C			
150 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
1	90	90	91
5	72	72	73
16	40	42	44
21	30	31	35

Group D			
100 ml water in each container			
Time (min)	White	Black	Metal
5	75	77	77
10	55	57	58
15	40	42	44
20	29	33	35
30	21	20	22

1. Decide which group you think did the best experiment. Explain why.

Continues

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2. Analyse the results from the group you think did the best experiment and decide which container has least heat loss (best keeps the temperature).

My conclusion is the

Black White Metal coloured They are the same

3. How certain are you about the conclusion? Tick one of the boxes below, or make your own statement.

I am absolutely certain that the container I identified is the right one.

The container I have identified is the best in the data, but the group's data gathering is too poor to make a final conclusion.

The container I have identified is the best in the data, but the results are very uncertain. Another test is likely to show different results.

It is impossible from the data to conclude if any container best keeps the temperature.

Other: _____

4. Compare your answers to the other students in your group. Decide between you what answers you think are correct.

Our final conclusion is that the _____ container(s)
best keep(s) the temperature, because

*Peer argumentation in the school science laboratory – exploring effects of task features***Tables and figures:**

	<i>Task Container</i>	<i>1: Task 2: Dissolving</i>	<i>Task 3 Discussion of Container task</i>
Group A (Girls)	x	x	x
Group B (Mixed)	x	x	x
Group C (Mixed)	x	x	x
Group D (Boys)	x	x	x
Group E (Mixed)			x
Group F (Boys)			x

Table 1: Design of the study

Level 1	Level 1 argumentation consists of arguments that are a simple claim versus a counter-claim or a claim versus a claim.
Level 2	Level 2 argumentation has arguments consisting of a claim versus a claim with either data, warrants, or backing but do not contain any rebuttals.
Level 3	Level 3 argumentation has arguments with a series of claims or counter-claims with either data, warrants, or backing with the occasional weak rebuttal.
Level 4	Level 4 argumentation shows arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counter-claims.
Level 5	Level 5 argumentation displays an extended argument with more than one rebuttal.

Table 2: Erduran et al.'s (2004: 928) Analytical Framework for Assessing Quality of Argumentation

Peer argumentation in the school science laboratory – exploring effects of task features

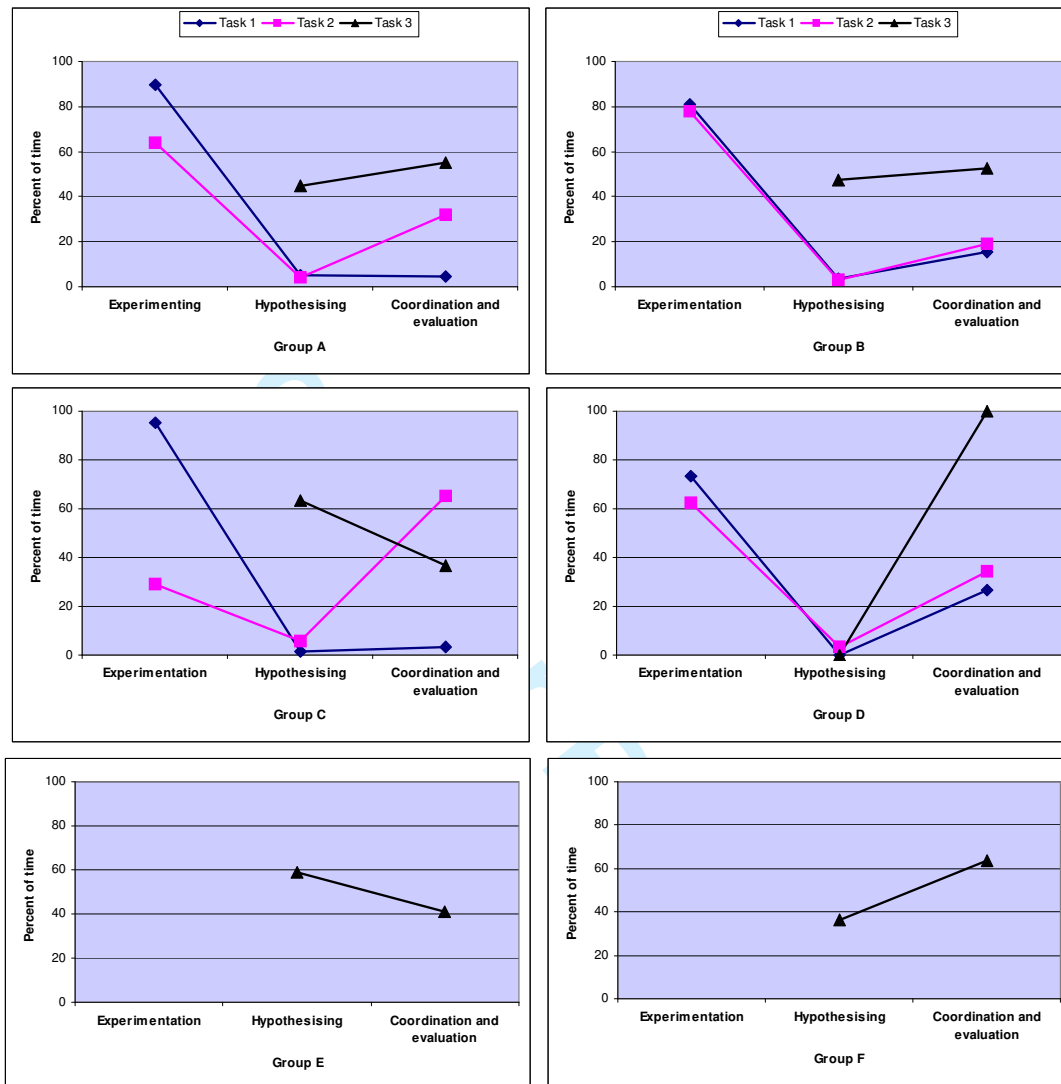


Figure 1: Orientation data: Total time spent by six student groups A-F on experimenting, hypothesising and co-ordinating and evaluating over three different tasks

Peer argumentation in the school science laboratory – exploring effects of task features

	Task 1	Task 2	Task 3	Total
Group A	7	10	8	25
Group B	4	7	6	17
Group C	9	17	3	29
Group D	3	6	1	10
Group E			11	11
Group F			2	2
Sum Gr A to D	23	40	18	81

Table 3: Number of argumentation units identified in each task each group

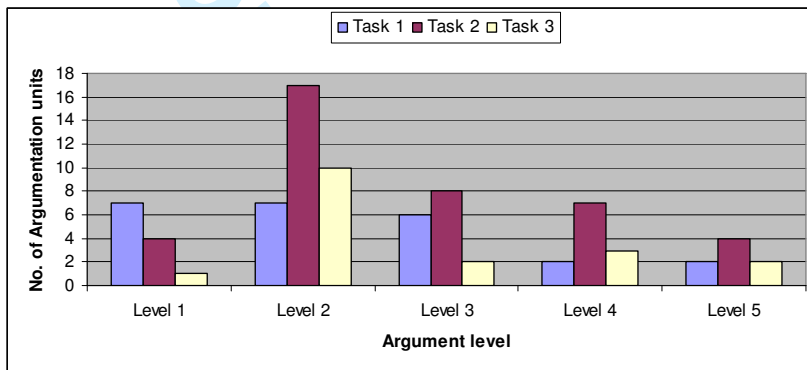


Figure 2: Number and level of argumentation units per task for Groups A to D.

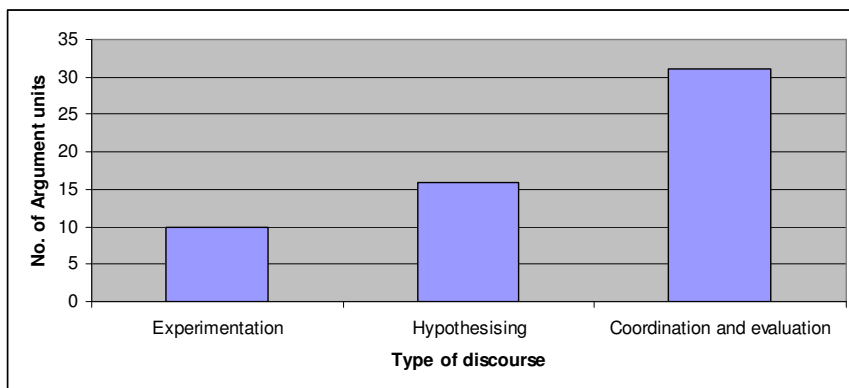
Peer argumentation in the school science laboratory – exploring effects of task features

Figure 3: Number of argumentation units summarised within each type of discourse for groups A to D.

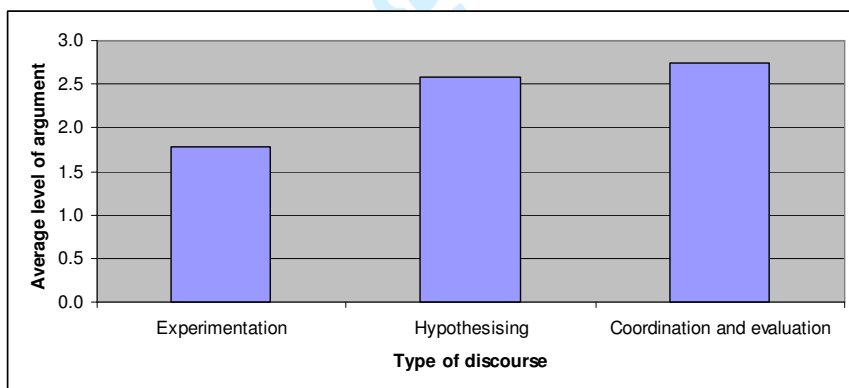


Figure 4: Average level of argumentation within each type of discourse for groups A to D.

Peer argumentation in the school science laboratory – exploring effects of task features

**Peer argumentation in the school science laboratory –
exploring effects of task features**

Prepared for *International Journal of Science Education*

By

Per Morten Kind
Durham University
School of Education
Durham, DH1 1TA, UK

Email: p.m.kind@durham.ac.uk

Vanessa Kind
Durham University
School of Education
Durham, DH1 1TA, UK

Email: Vanessa.kind@durham.ac.uk

Avi Hofstein
Weizmann Institute
Rehovot Israel

Janine Wilson
c/o School of Education
Durham University
Durham DH1 1TA, UK

*Peer argumentation in the school science laboratory – exploring effects of task features***Abstract**

Argumentation is believed to be a significant component of scientific inquiry: introducing these skills into laboratory work may be regarded as a goal for developing practical work in school science. This study explored the impact on the quality of argumentation among 12 – 13 year old students undertaking three different designs of laboratory-based task. The tasks involved students collecting and making sense of complex data; collecting data to address conflicting hypotheses; and, in a paper-based activity, discussing pre-collected data about an experiment. Significant differences in the quality of argumentation prompted by the tasks were apparent. The paper-based task generated the most argumentation units per unit time. Where students carried out an experiment, argumentation was often brief, as reliance on their data was paramount. Measurements were given credence by frequency and regularity of collection, while possibilities for error were ignored. These data point to changes to existing practices being required in order to achieve authentic, argumentation-based scientific inquiry in school laboratory work.

*Peer argumentation in the school science laboratory – exploring effects of task features***Introduction**

A rapidly growing interest in *argumentation* among science educators has been fuelled by socio-constructivist theoretical frameworks in science philosophy and learning psychology (Jiménez-Aleixandre & Erduran, 2008). In science philosophy, science theories are no longer seen as products of pure empirical or logical processes but as ideas shaped through critique, debate and revision within the science community (Kuhn, 1962). In learning psychology, learning is seen as originating from socially mediated activities (Vygotsky, 1978). Knowledge and cognitive processes exist in a social milieu and are internalised by the individual through language and active participation. From these key ideas it has been claimed that science education needs to change from a current “positivist” practice emphasising a misleading picture of science and factual recall of knowledge, towards a practice that helps socialisation of young people into the norms and practices of authentic science (Driver, Newton, & Osborne, 2000). The current paper relates these changes to teaching and learning in the science laboratory. To date, socio-scientific issues have provided a more common focus for argumentation in science education activities, perhaps because preparation for decision-making in political and moral issues students meet in society (Zeidler, Sadler, Simmons, & Howes, 2005) has dominated educational thinking. Laboratory activities may be an attractive alternative drawing the attention towards argumentation in a scientific context and demonstrating something about the architecture of scientific knowledge (Ford, 2008). Osborne, Erduran and Simon (2004) offer a second reason for the dominance of the socio-scientific context, by showing that initiating argument in a scientific context is much harder. To keep argumentation meaningful students need to understand evidence (Koslowski, 1996), which can be challenging for some scientific phenomena. In this view, the laboratory also may be favourable because students familiarise themselves with the phenomenon and generate their own data.

There are, however, obvious reasons that argumentation in the laboratory is difficult (Hofstein, Kipnis, & Kind, 2008). A typical school inquiry more closely resembles the empiricist-logical view of science that dominated science philosophy a century ago than today’s socio-constructivist view (Driver et al., 2000). Scientific method is presented in a step-wise manner leading students steadily along a path from research problem to final conclusion. Thus, they are led in ways that avoid the complex nature of authentic science to ensure that the “right” data appear and “correct” conclusions are made. Open-ended investigations with more degrees of freedom occur, but are limited in their use due to the pedagogical challenges they present to teachers (Hofstein & Lunetta, 2004). Meeting curriculum targets and conforming to assessment practices are also restrictions (Donnelly, Buchan, Jenkins, Laws, & Welford, 1996). Establishing the science laboratory as an efficient place for teaching and learning scientific argumentation requires that we face the challenge of breaking these practices.

Against this background the current paper presents an exploratory study that has investigated the effect of specific pedagogical strategies in laboratory tasks. Chinn and Malhotra (2002a) analysed laboratory task formats extensively. They define one group of

Peer argumentation in the school science laboratory – exploring effects of task features

tasks as “simple experiments”, the main focus in this study. These are typically tasks in which students research the effect(s) of one or two provided variables, with some openness in choice of method. In the UK, the locus for this study, such tasks are referred to as *investigations*. By suggesting particular pedagogical strategies implemented in the tasks, trialling these in “ordinary” teaching contexts and accepting that classroom life is synergetic (Brown, 1992), we hope to shed light on why argumentation in the science laboratory presents challenges and suggest specific actions that may be taken to improve current practices. Before introducing the study the paper reviews research analysing discourse practices in the science laboratory and sets out the underlying rationale.

Argumentation in the school science laboratory

Science education research literature portrays the laboratory as a complex environment for teaching and learning (Harmon, Smith, Martin, Kelly, Beaton, Mullis, Gonzalez and Orpwood, 1997; Hegarty-Hazel, 1990; Hodson, 1993; Lunetta, Hofstein, & Clough, 2007; Séré, Leach, Niedderer, Psillos, Tiberghien & Vicentini, 1998). On one hand, it seems straightforward, familiarising students to do investigative work. For example, the Assessment for Performance Unit (APU), a large UK-based assessment project, collected data from students doing practical investigations in the early 1980s (Schofield, Black, Head & Murphy, 1985; Welford, Bell, Davey, Gamble, & Gott, 1986). Researchers were positively surprised about the relatively advanced levels of student performance (Murphy & Gott, 1984). Despite students having limited experience of carrying out investigations, they approached the APU tasks with ‘confidence and success’ (p 40). Around half of the students were judged to carry out ‘good’ experiments, with only a fifth or less failing to make the expected quantitative measurements. Similar positive data are reflected in the Third International Mathematics and Science Study (TIMSS, Harmon et al., 1997). On the other hand, getting students from a basic level to using the laboratory for more sophisticated learning is difficult. The APU studies (op cit) showed surprisingly little progress when 13 and 15 year-old students were compared. Students seem to learn elementary ways of solving laboratory tasks quickly, but do not progress beyond these. One contributory factor is the standardised form of laboratory tasks, which invite students to learn ‘scripts’ (Schank & Abelson, 1977). Many will have observed with Millar, Lubben, Gott & Duggan (1994) that students start solving investigative tasks without first spending time on planning what to do. The wider meaning, of course, is that the science laboratory has become a place for conducting routine exercises (Lunetta, Hofstein & Clough, 2007) so that students are seldom challenged to reflect on the methods they use and/or observations they make (Hodson, 1996). Recent evidence confirms that this problem still applies to UK school science laboratory teaching. Abrahams and Millar (2008) observed twenty-five randomly selected laboratory lessons and found that teaching *seemed* well-functioning on the surface, as students gathered data and conducted activities as their teacher expected. However, students rarely reflected on their findings and methodology. Unsurprisingly, Newton, Driver and Osborne (1999) document the absence of argumentation in ordinary laboratory teaching. As little as 0.4% of time spent on laboratory tasks was organised as group discussion. In fact, laboratory lessons included *less* discussion than non-laboratory lessons. Teacher questionnaires and twenty-three case studies across Europe suggest a similar trend in other countries (Séré et al., 1998). Noticeably, even if students work in pairs or small groups, the most common

Peer argumentation in the school science laboratory – exploring effects of task features

organisation for laboratory work, the overall picture is unchanged. Laboratory group work generally means students working alongside each other following teacher and task instructions, not conducting *group discussions*.

Despite this situation, scientific argumentation could be taught efficiently in the laboratory. Importantly, the argument is commonly turned the other way around, claiming that *if* students are doing investigations in the laboratory these *should* include argumentation; because, unless they do, the activities portray a misleading image of science (Driver et al., 2000). From this view, however, we deduce that training students to do investigations could be training in scientific argumentation. Hence, this rationale underpins intervention studies exploring ways of teaching scientific argumentation (Cavagnetto, 2010). Nevertheless, few studies utilise investigation contexts in which students collect their own data in a laboratory setting. Alternatives such as computer-based experiments (Clark & Sampson, 2008), observations from video (Engel & Conant, 2002) and analysis of a secondary data set (Sandoval & Reiser, 2004) are apparent. A common point, however, among all these studies *and* those using laboratories is that the investigation provides *scaffolding* for argumentation teaching. In other words, when teaching argumentation through scientific investigation, teaching can be structured around debating alternative hypotheses, analysing data and evaluating evidence for a conclusion. This, we may claim, strengthens the case for teaching argumentation in the laboratory compared to many other contexts, because the scaffolding is already in place. As we have seen, problems occur when students mechanically follow ‘scripts’ in their laboratory investigations, but breaking out of this by ‘forcing’ students to reflect on and debate what they do and observe may provide a good way forward. Several intervention studies follow this line of argument. For example, Watson, Swain and McRobbie (2004), conducted an open-ended laboratory task running over three lessons in a UK school. They stimulated argumentation by guiding students towards ‘focusing’, ‘planning’, ‘obtaining evidence’, ‘interpreting’ and ‘evaluating’. In Israel, Katchevich, Mamlok-Naaman and Hofstein (2010) allowed students more time to plan open-ended investigations and encouraged discussion at various stages during their work. Both studies report positive outcomes compared to the ‘ordinary’ laboratory work Newton et al. (1999) describe. However, although these studies increased the amount of argumentation in laboratory investigations, both found the *quality* of argumentation to be relatively poor. Watson et al. (op cit) found students mostly met claims with counter-claims rather than using data from their investigation. Katchevich et al. (op cit) similarly report low level argumentation as measured by a coding framework. Two other US studies suggest improved argumentation quality depends on persistent use over a longer period (Richmond & Striley, 1996; Yerrick, 2000). Yerrick’s study was carried out over eighteen months. Pre- and post-intervention interview data reveal that the number of times students linked observational evidence with their proposed warrants more than doubled over this time period. Richmond and Striley’s students did four investigations over a three month period. All data gathering was qualitative, but in their conclusions they state:

One of the most significant changes we observed over the course of these four experiments was in students’ ability to formulate appropriate scientific arguments: They became more adept at identifying the relevant problem, collecting useful information,

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

stating a testable hypothesis, collecting and summarizing data, and discussing the meaning of data.

(Richmond and Striley, 1996: 847).

Besides offering structure for argumentation teaching, benefits may also accrue in the laboratory from the availability of students' data, which act as a stimulus for debate. Any teaching activity aiming to create debate depends on stimuli to generate different opinions and contrasting assertions. Activities placed in a socio-scientific context, for example, commonly utilise political and ethical controversies (Sadler, 2004). Students are stimulated to engage in debate by taking and defending a stand. Cognitive conflicts or misconception-based discussion can also act as stimuli (e.g. Naylor, Keogh, & Downing, 2007). In student-to-student discussions *one* authoritative voice is less likely and debate is triggered because different conceptions are 'compete'. Data interpretation is a common stimulus for argumentation (Sandoval & Millwood, 2007). Debate is created about different hypotheses arising from the data and their value as evidence for stated conclusions. This, however, may work better when data are collected by students themselves rather than taken from other sources, because this generates students' sense of ownership. No study has actually tested this potential effect, but Simonneaux (2001) and Jiménez-Aleixandre, Bugallo Rodriguez & Duschl (2000) suggest that personal involvement with both data and arguments is important.

Thus, we see that scientific investigations conducted by students offer a *natural locus* in science education for teaching argumentation. Firstly, the investigation manifests what we mean by scientific argumentation; secondly, it provides a scaffold for teaching; and, thirdly, it creates ownership to arguments by being based on students' own data. Katchevich, Mamlok-Naaman, & Hofstein, (2010), Richmond & Striley (1996), Watson et al. (2004) and Yerrick (2000) all attempt to improve argumentation in the science laboratory by using strategies strengthening these aspects. Internationally, other intervention studies, however, suggest that teaching should go further, and do more to 'stage' debate. In Korea, for example, Kim and Song (2006) stimulated argumentation in an open-ended investigation by having students present reports for 'peer review', and arranging discussions in which students acted as critics in a similar manner to scientists attending a conference. In Norway, Kolstø and Mestad (2005) used peer reviewing of laboratory reports as teaching approach. In the UK, Naylor, Keogh and Downing (2007) used concept cartoons to create 'conflicts' before the investigation and a whole-class session to debate findings afterwards. In the US, Martin and Hand (2009) combined investigations with presenting and analysing claims in writing. These research studies are not presented to permit precise comparisons of the exact amounts and/or quality of argumentation. Nevertheless they offer the impression that these strategies are important supplements to 'ordinary' student investigations. A pertinent question, however, is to what extent such additional strategies can be sustained in laboratory teaching? Some strategies span laboratory activity over several lessons and restrict the number of activities students can do within a school year. They also rely on science teachers feeling comfortable about participation in this type of activity.

Peer argumentation in the school science laboratory – exploring effects of task features

In the current study we isolate and explore effects of using various strategies when teaching argumentation in the school science laboratory. We built on traditions established in science education and stayed within the range of teaching approaches normally used by science teachers. The main focus was to explore issues of achieving better scaffolding of argumentation during an investigation and to probe how different forms of data stimulate debate.

Classroom cultures and the role of the teacher

Learning science is learning to talk science (Lemke, 1990). The specialist language of science not only includes declarative and procedural concepts, but also carries an ideological position with reference to certain epistemological criteria (Kelly, 2007). The process of becoming a scientist involves adapting to a scientific way of thinking and working by participating in science culture over time. Although few students become scientists, a similar process of enculturation is put forward as an ideal for any science classroom (Driver et al., 2000). Science teaching should offer a learning environment that promotes the values and norms of science and gives students rich opportunities to engage in scientific discourse. The typical discourse of the science classroom, however, is tightly controlled by the teacher, using what Lemke (op cit) refers to as the ‘triadic dialogue’; a question-answer-response pattern that gives little room for students to talk science and practice making the language of science their own. Many have therefore concluded that one of the biggest challenges of teaching scientific argumentation is to create a classroom environment less dominated by teachers holding on to a didactic teaching style (Herrenkol, Palincsar, DeWater, & Kawasaki, 1999; Jiménez-Aleixandre & Erduran, 2008; Kuhn, 1992; Zohar & Nemet, 2002).

Ideally, the teacher should fulfil two roles. One is *facilitator* encouraging students to participate in discussions by presenting views and critiquing those of others (Simon, Erduran & Osborne, 2006). When applied to investigative work, this means encouraging students to present and discuss alternative hypotheses and alternative ways of solving tasks, rather than checking if they have the ‘correct’ solution. Students need to feel confident and able to accept that disputes are accepted and natural to scientific inquiry. The second role is to model good scientific practice. Jiménez-Aleixandre (2008) describes this as being an *able peer*, providing and persuading the use of scientific epistemic criteria. This can be done by engaging in debate and asking open questions aimed at eliciting justifications (Jimenez-Aleixandre, Lopez Rodriguez, & Erduran, 2005). The teacher may challenge students’ ideas, pointing out limitations and inconsistencies (Mork, 2005). These two roles plead for an active teacher engaging as a ‘partner’ to student groups running their own debates.

The teacher’s importance, of course, may undermine the relevance of exploring strategies in investigative tasks. Firstly, if the classroom does not have an environment supporting argumentation, no teaching strategy is likely to prove efficient. Secondly, from a research methodological point of view, the classroom environment may be a confounding variable seriously distorting what we want to investigate. To compensate, the current study emphasised creating a learning environment in accordance with the ideals presented above and keeping this constant in all teaching.

*Peer argumentation in the school science laboratory – exploring effects of task features***The study**

The study started by identifying common investigative tasks relevant to the England and Wales science National Curriculum (Qualifications and Curriculum Authority, 2007). Although variation exists, some tasks are used repeatedly in many classrooms, partly due to guidance from assessment authorities (Donnelly et al., 1996) and partly for practical reasons: finding good research or genuinely open-ended problems suitable for student investigations in school contexts is thought to be difficult. Good examples are easily established through science course materials, so tend to reoccur often in lessons. Typically, tasks require students to find the effects of one or more independent variables on a single dependent variable. Even if tasks are characterised as “open-ended”, students are often given the problem and follow standardised, “script”-based procedures to arrive at pre-determined solutions. Newton et al. (1999) observed such “open investigations” being used in four out of twenty-three randomly observed practical work lessons. Two typical tasks were chosen for this study (see below), selected partly because they are familiar to teachers but also because the equipment needed is very simple.

Investigations are at the heart of the problem of creating argumentation in the science laboratory: they are intended to introduce students to methods of scientific inquiry, but often fail to do so because of practical and pedagogical constraints (Chinn & Malhotra, 2002a). These activities involve no discussion about the evidence generated, or the way a conclusion has been drawn, since data are “unproblematic” and the answer “obvious”. Our next step was therefore to reflect on ways in which this pattern could be broken, changing tasks to prompt more authentic scientific argumentation. Working from the rationale suggested by the introduction and research studies promoting argumentation strategies in science lessons more generally (e.g. Osborne, Erduran & Simon, 2004) we identified the three strategies outlined below.

Use of complex data

This involves leading students into a situation where data do not give one obvious “correct” conclusion. We adapted a task investigating the temperature drop occurring over time when containers of different materials are filled with hot water (Appendix 1, Task 1, “*Container*”) for this. The activity was set in the context of investigating transportation of hot liquid in metal tankers with different surfaces. Students were provided with metal containers (empty, clean food tins) coloured black, white and (shiny) grey metal. The material was identical, so only the surface colour varied. Small differences in temperature drop occur due to heat radiation. The shiny container exhibits least radiation and consequently the smallest temperature drop. As all temperature drops over time are small, students produce measurements in which differences in data collected from the three containers are of similar magnitude as the level of uncertainty in each measurement.

We anticipated that students would be led into and take different stances in a debate about data as evidence for claiming which container best retained the water temperature.

Peer argumentation in the school science laboratory – exploring effects of task features

Conflicting hypotheses

Here we presented students with conflicting theoretical hypotheses, about which they were likely to disagree. The students debated which of a set of statements were right or wrong *before* starting the investigation, and then to investigate the truth of these by experiment. Dissolving salt (sodium chloride) in water of different temperatures (Appendix 2, Task 2, “*Dissolving Salt*”) was the task. Initial statements were based on the common misconception among 11 – 16 year olds that sugar and salt lose mass when dissolved in water (Ebenezer & Erickson, 1996).

We anticipated that students would support different hypotheses, or, as a minimum, perceive the investigation as testing competing hypotheses, and engage in debate about these during phases of the investigation.

Post-investigation discussion

The third alternative differs from the first two in that a discussion is arranged *after* the practical investigation is complete. Handling equipment and gathering data are complicated tasks, so arranging a post-experiment discussion is an alternative to ‘scaffolding’ that may prompt students’ reflecting about scientific inquiry. A task was developed that followed the “*Container*” (Task 1) practical investigation described above. Hence, this third task shared the element of using data to stimulate argumentation with “*Container*”. Students were presented with fictional data related to “*Container*” collected by imaginary student groups who had used different methods (Appendix 3, Task 3 Post discussion). Questions directed students to evaluate the methods the data produced, and to debate evidence for the conclusion. Students answered questions individually and then in group debated in order to reach a common conclusion.

We anticipated that students would take a more holistic perspective on the investigation, looking at the methods for data gathering and the results when debating the final conclusion to be drawn.

The study investigated the effects of each alternative task on 12 - 14 year old students’ argumentation when working in small groups. The research questions probed were:-

1. To what extent do these alternatives stimulate scientific argumentation?
2. If argumentation occurs, what forms of debate arise?

Design and methodology

The design was based on testing the three different strategies in ordinary whole class laboratory teaching. The tasks were planned for lessons lasting 60 minutes (a lesson length common in secondary schools in England and Wales), at an independent school (i.e. private, non-state funded) in the North East of England. The participants were three Year 8 (12-14 year old) mixed ability classes, each with 22-23 students. The science teacher (author 4 of the paper) taught all classes. This ensured consistency across the teaching situations and a classroom climate that encouraged argumentation as indicated above.

Peer argumentation in the school science laboratory – exploring effects of task features

Two classes carried out all three tasks in small groups. From these classes, four groups, each of three students, were selected for in-depth analysis. From the third class two groups were selected. These acted as “controls” by doing only Task 3, the “post-investigation” that is without first having conducted the matching practical task (1). Groups were sampled to mix gender and ability. Table 1 shows the data gathering design. The design made it possible to compare different student groups doing the same activity to see how a strategy worked for four different student groups. We were also able to compare the same group doing three different tasks to see how a student group reacted to three different strategies. The two groups doing only Task 3 enabled comparison with students who did this task with and without carrying out the practical task (Task 1, *Container*) first. Data were gathered over a two week period.

(Table 1 in about here)

Data gathering was made by video recording using one high quality Sony DVCAM camera on each focus group, and one wireless “fly” microphone attached to each student. This enabled detailed recordings of student activities and their conversation in a classroom with background noise. The microphones also recorded conversations between focus group students and the teacher. Using video camera and microphones, of course, influences students’ talk, but students in all groups kept talking freely and their behaviour suggested no obvious constraints.

Students produced written reports, one per group, from each task. These were collected and used as background information to support and validate video recording data. In Task 3 students also produced written responses for comparison.

The observed lessons began with 5 – 10 minutes teacher introduction to the task. Emphasis was placed on students being explicit about their thinking and needing to make decisions during the investigations. They were also told that conclusions should be evidence-based and that achieving this should be a particular focus. The students were used to working in the laboratory and familiar with investigative tasks. No specific teaching, however, had been provided on argumentation. The introduction and clearing up afterwards left 30 to 40 minutes for each practical investigation (Task 1 and 2). Task 3 required less time and lasted 10 to 20 minutes, depending on how quickly students handled the task.

Data analysis

Data analysis started with first transcribing all conversations from the video recordings. The transcripts were checked by a second, qualified academic colleague to ensure their correctness. The remaining data analysis was made from the transcripts, but often with a return to the video recordings to have a better understanding of the context or by confronting students’ written reports. Although this is a small scale study, the analysis was both qualitative and quantitative.

*Peer argumentation in the school science laboratory – exploring effects of task features****Identifying argumentation discourse***

The main focus of the analysis was to identify the type of argumentation discourse occurring in the student groups, but for comparison between tasks we also “measured” frequency and quality of argumentation. Several possible analysis methods were considered, as there is no one obvious method (Erduran, 2008). Watson, Swain and McRobbie (2004) counted claims made by students and looked at how often these were supported by data. Newton et al. (1999) focused more generally on the orientation of students’ work, e.g. “group discussion” and “closed experiments”, and measured time devoted to each of these. Students’ discussion in the different orientations were then studied qualitatively, and together with the quantitative time measure gave an indication of the amount of argumentation happening. A third approach, adopted here, is to identify “argumentation units” using Toulmin’s (1958) Argumentation Pattern (TAP). TAP describes argumentation with a claim being supported by data, and with various ways of strengthening or undermining this relationship. Each unit, of course, can last for various lengths of time, but this is subordinate to the quality of the argumentation contained. The argumentation measure is therefore given by number of units (more units indicate more frequent argumentative discourse) and the quality of each. Counting and scoring the units is not straightforward, because of the difficulty related to deciding what is claim, data and warrant (Erduran, 2008). When a new claim is made, for example, it may sometimes be judge as a simple “counter claim” (meeting a claim with another claim), and therefore belong to the same unit, or be seen as the start of a new unit. A depending factor is how it is supported with data and warrants. The numbers produced for units should therefore be read with some caution. They give meaning mainly when compared within the study: here, the researchers have been able to compare units and agree on definitions used consistently in the three tasks. In accordance with Zohar and Nemet (2002), claims made without any justification or not met with a counter claim were not recognised as an argumentation unit.

Determining the quality of argumentation discourse

The criteria used for scoring quality of argumentation units relates to low quality argumentation being “sparse” with few backing or rebutting elements and high quality argument “rich” in such elements. To score this issue we used a classification system developed by Erduran, Simon and Osborne (2004, see Table 2). Other frameworks were considered (an overview is presented in Sampson & Clark, 2008), but Erduran et al.’s (op cit) framework was selected because the five level scale provides a means of grading students’ comments. The key discerning factor is the presence or absence of rebuttals. In the two lowest levels there is no questioning of claims. Students either meet a claim with another claim (Level 1) or use some form of argument to support their claim (Level 2). The three next levels considered the quality and quantity of rebuttals. Identifying and coding argumentation units were equally important for the qualitative analysis of the group discussions as it was for the “scoring” and “counting”. The close inspection of units served as a means for understanding the nature of the laboratory discourse.

(Table 2 in about here)

Peer argumentation in the school science laboratory – exploring effects of task features

Two researchers identified units and carried out the coding for each piece of transcript. When units were agreed, inter-coder reliability was 70 - 80 % when coding their levels. Disparities were resolved through discussions.

Student orientation

An additional coding of students' "orientation", or focus, while working on the tasks was also conducted. This coding was not planned in advance but derived from the data as an attempt to characterise ways in which students solved the investigation tasks. Theoretically the coding has support in Klahr, Fay and Dunbar's (1993) *Scientific Discovery as Dual Search* (SDDS) model. This model suggests that someone (students and scientists) working on an inquiry task may operate in different "problem spaces". The model has two problem spaces, the *experimentation space* and the *hypothesis space*, but Klahr et al. add a third dimension, making in total three orientations:

- a) Experimentation: Students are focused on data gathering and handling of equipment
- b) Hypothesising: Students are focused on explaining the observed phenomena by use of scientific theories and concepts
- c) Co-ordination and evaluation: Students are focused on co-ordination and evaluation of evidence to draw a conclusion.

By using the SDDS model as a theoretical underpinning we suggest that students may be operating in these orientations disjointedly and with some problems combining them or moving from one orientation to another. Further details about how the coding was conducted will be given in the result section below.

Results

We will start by presenting two examples demonstrating students' complete work on Task 1 and 2. The examples illustrate characteristics of group work being "rich" (No 1) and "poor" (No.2) on argumentation. As such, these represent extremes in the findings. Number 1 is Group A carrying out Task 1 (*Container*, complex data) and Number 2 is Group C carrying out task 2 (*Dissolving Salt*, conflicting hypotheses). Students in the groups have been given fictitious names with first letters matching the group letter. Student comments are reported verbatim with editorial additions shown in square brackets to ensure the context is clear.

Example 1: Group A on Task 1, "Container" with complex data

After having been informed about the task and organised as a group the three girls in Group A went directly on to collecting equipment and setting up the experiment. No discussion occurred about the purpose of the task. They exchanged single comments as they went along with the preparations, repeating information from the task sheet or given by the teacher, to clarify and agree what they were doing. For example (brackets show time in minutes: seconds):

Ann (0:25): We need to explain clearly what goes into the container.

Peer argumentation in the school science laboratory – exploring effects of task features

Nearly nine minutes later, three metal containers with different surfaces (white, black and metal), each with a thermometer were lined up at the table and filled with hot water from a kettle. Then came explicit comments about the research question:

Amy (8:50): We need to do something..

Ann (9:03): We need to see which one ... like keeps the most heat..

It was clear that the students' focus so far had been *what* to do rather than *why*. Having made their first recordings of temperature in each container the discussion continued:

Ann (9:22): Do they [the company mentioned in the task] want to keep it hot .. or do they want to keep it cold?

Ada (9:29): I don't really know

Amy (9:29): They want to keep it warm

Ann (9:30): They want to keep it hot, okay

Ann (9:31): Okay, so we see which stays the hottest

Ali (9:32): This is the hottest at the moment...

This was their only explicit "planning" and in the following twelve minutes the students were fully engaged with recording data. As they went along, however, three more methodological issues (*italicised*) were brought up, all in the same accidental way and each was solved in a single sentence:

First, *how to present data*:

Amy (11:31): So.. we like, make a table

Ada (11:31): [Grunts] Yeah

Second, *what was the temperature of the water when they started?* (They forgot to record temperatures immediately after having poured the water.)

Ada (13:15): What was the temperature at zero, Amy?

Amy (13:17): Temperature at zero was 100 degrees for everything, because it had all come out of the ..eh.. kettle.

Third, *how long should they keep taking measurements?*

Amy (21:15): How many minutes do we have to do it for?

Ann (21:17): I don't know

Amy (21:20): Just keep going until she (the teacher) stops us

Ann (21:24): Yeah

During the data gathering analysis was also carried out in a similar way, as the girls started to discuss which container best keeps the temperature. This started with Amy

Peer argumentation in the school science laboratory – exploring effects of task features

making a comment after the ninth set of readings, in which the temperature in the metal coloured container (“silver”) was 86 °C, the black was 84 °C and the white was 81 °C:

Amy (18:47): I think silver is going to be the best, don't you

Ada (18:49): Yeah

A few minutes later Amy declared the “winning container” and told the teacher they had finished the investigation. The teacher wanted the students to evaluate the data before drawing a conclusion, but the students did not see a need for this. When the teacher asked the students *why* the metal surfaced container had best kept the temperature, it became clear that the students had expected the black container “to win”:

Ada (24:58): [Be]cause it absorbs heat, like when the Sun is...

Amy (24:58): Like when you wear black clothes in the Sun

Ada (24:58): Yeah, in the Sun and then it absorbs the heat

These ideas, however, were never discussed among the students while doing the investigation, or in the students' report, which just presents the measurements and confirms that the metal coloured container came out with the smallest fall in temperature.

Example 2: Group C, Task 2: “Dissolving Salt” with conflicting hypotheses

The students in Group C, two boys and a girl, started by reading out the three statements in the task (see Appendix 2) and discussing which were “right” or “wrong”. Although lasting less than two minutes the discussion made a basis for the further investigation. Two students (boys Cameron and Callum) supported the statement “mass cannot disappear” and claimed the opposite statement “there is always a small loss of mass” therefore had to be wrong. The third (female) student, Cynthia, was less sure and thought statement B might be right. Statement C (mass depends on the temperature of the water) was declared wrong by Callum. The two other students expressed some doubts about this, but Callum's self-assuredness silenced their views. The final outcome, decided mainly by Callum, was to do experiments to test statements A and B.

The students decided on using 25ml water and 1 g salt to test hypothesis A. These were measured independently before the salt was dissolved in the water and a new measurement made of the solution. With small amounts, the measurements were not exactly the same before and after dissolving the salt. There seemed to have been a loss of mass, which confirmed Cynthia's view:

Cameron (16:15:): Some is gone... B is right, I think

Cynthia (16:16): I said B. Didn't I say B!

At first they were all convinced about this conclusion, even Callum:

Callum (16:32): So there is always a small loss of mass when it dissolves...

The mass will increase, but less than the mass of the salt added.

Peer argumentation in the school science laboratory – exploring effects of task features

Cameron (16:47): That is right.

Callum (16:47): So B is right.

Cynthia then suggested they had *disproved* statement A only and that they still had to *prove* statement B. A 15 minute discussion started in which Callum tried to convince the others that having a test disproving A is the same as proving B to be right. The outcome of the discussion was an agreement that they should do the same test again. When doing this they tried to copy the exact amounts used in the first experiment, arguing that “things had to be kept the same”. This, however, was difficult and their next set of readings was different:

Callum (21:26): Shouldn't it be the same as that [compares measurements from the two experiments]

Cynthia (21:30): No, it wasn't, like, precise. You can't get ... or you can, but it would be really hard to get exactly the same amount

This comment from Cynthia gave Callum a sudden understanding that hypothesis A still might be right and that the unexpected results in the first experiment was due to measurement error:

Callum (21:36): No, wait a moment (eager). That means that that [statement A] could be right. Because, it is really hard to get the exact measurements.

He continued:

Callum (21:56): I do not think we lose anything. It is just impossible to make it precise.

Callum tried to convince the others about this. They, however, still thought they should stay with the outcome of the measurements, which indicated a loss of mass (i.e. supported statement B). The discussion included detailed analysis of the experiment and why or why not the final outcome should be trusted. Callum did not manage to convince the others on his view.

Patterns of orientation in students' investigations

The two examples show different approaches to the investigations. Example 1 demonstrates an algorithmic approach, focusing on data gathering to answer a question. The girls implicitly “knew” what to do, not needing planning, and they unreservedly trusted the outcome of the measurement, not perceiving a need to evaluate. A conclusion could be drawn, and the investigation was “finished” as soon as the last measurement was made. Example 2 developed differently. More time was spent on planning and on conceptual discussion, but the most characteristic difference from Example 1 is extensive discussion about what conclusion could be drawn from the data. This was absent in the first group, but filled more than half the time of the second group. The difference has an obvious importance for occurrence of scientific argumentation.

Peer argumentation in the school science laboratory – exploring effects of task features

As a way of demonstrating these differences across all groups and tasks a decision was made to code time spent on the different orientations of students' discourse. Social talk, talk about the task (rather than problems in the task) and discussing/writing the report were excluded from this analysis and coded as "other", as these were not focii. The transcript was the main source of coding, with support from the video pictures. Sequences of the group investigation were coded: a) Experimentation, b) Hypothesising or c) Co-ordination and evaluation, and the total time spent on each noted.

Figure 1 shows how time was spent by the groups on each orientation over the tasks they carried out shown Table 1. The category "other" is excluded, so the percentage is therefore a relative distribution of time spent between the three main categories.

[INSERT Figure 1 about here]

The graphs show that Task 1, *Container* with complex data, prompted students to focus on "experimentation". Group A (see above), held this focus 90% of the time, hardly devoting any time to "hypothesising" and "coordination and evaluation". Group C on Task 2, *Dissolving Salt*, conflicting hypotheses, (see above) is an exception, spending more time on coordination and evaluation than on experimentation. Task 2 stimulated more discussion about co-ordination and evaluation than Task 1. However, we sense some "randomness" in the data: how a task develops for the groups is inconsistent. Groups B and D exhibit similar patterns for Tasks 1 and 2, indicating that orientation is group-related (groups having the same orientation across tasks), while the opposite is the case for Groups A and C, indicating that orientation is task-related. All in all, this indicates a situated effect, suggesting that the way in which a task develops for a group is influenced by factors happening during that particular event.

Task 3, the post-investigation discussion, naturally has a higher percentage of the time on the two last categories, since it was non-practical. We see that students, except for Group D, balance time between "hypothesising" and "coordination and evaluation". This indicates that hypothesising, obviously very difficult to stimulate during the practical work, was more easily stimulated in this task format.

Argumentation and students' orientation

Table 3 summarises the number of argumentation units in each of the three tasks. We see that Task 2, *Container*, had more units than the two other tasks, generated by groups A and C. Task 3, the post-investigation discussion of *Container*, had fewest units. It is, however, important to keep in mind that Task 2 lasted 30-40 minutes while Task 3 was solved within 10 to 20 minutes

(Table 3 about here)

Figure 2 presents the level of the argumentation units in Table 3 scored using Erduran et al.'s (2004) framework. We judged most argumentation at level 2, in which students

Peer argumentation in the school science laboratory – exploring effects of task features

present a claim with some form of justification, but without rebuttals. Tasks 2 had more units at higher levels than the other tasks, stimulate by ‘conflicts’ established in the pre-discussion. In discussions trying to solve these conflicts later in the investigation (as demonstrated in Example 2 above) students made use of data and presented rebuttals. This brought the argumentation to higher levels.

(Figure 2 in about here)

Figure 3 presents the number of argumentation units in each of the three “orientations” presented earlier, but summed up across all tasks. 12 % only (10 out of 81) happened in “experimentation”, which took on average more than eighty percent of students’ time in the practical tasks. Argumentation occurred most frequently when students were coordinating and evaluating the evidence and conclusions.

(Figure 3 in about here)

Figure 4 suggests that argumentation unit quality was enhanced to 2.7 on average when students were “hypothesising” and doing “coordination and evaluation”, compared to 1.8 for “experimenting”. Argumentation in the “experimentation” mode meant mainly students making claims in relation to the data recording. For example, in Task 1, claiming one container to be “best” and supporting this with temperature measurements. A claim like this was unlikely to be rebutted by another student, so the argument unit was very short. Some discussion occurred about methods, triggering some more advanced structures, but this was rare and also relatively short. For example, Darren in Group D claimed their experiment on dissolving salt (Task 2) was “false”:

Darren (12:03): ..we have changed the amount of water since the first time we weighed it, so this mass will be totally different. So this experiment is false.

Some discussion developed from this, but the problem was solved by carrying out the experiment again. The argumentation occurring when “hypothesising” had higher levels because students rebutted each others’ explanations, but this again was not very frequent. In “coordination and evaluation” mode, the type of argumentation had much more variation. Students could go into detailed discussion about data in order to solve disputes about what conclusions to draw. This has already been demonstrated by the example of Group C solving Task 2, but a more detailed illustration of the type of argument provided by Group D’s discussion of the statement “mass is dependent on temperature of the water” in the same task. Darren claimed temperature has no effect, but was challenged with the rebuttal that “salt was at first observed at the bottom of the beaker, but disappeared when the temperature increased”; based on the warrant that solid salt and dissolved salt do now weigh the same. Darren first agrees with the observation but attacks the warrant:

Darren (24:00): The higher the temperature the more salt can dissolve, but it does not affect the mass; unless you put more salt in. If

Peer argumentation in the school science laboratory – exploring effects of task features

you have the same amount of salt and the same amount of water, heating doesn't actually change the actual mass that is in there

He then strengthens his own argument with a rebuttal

Darren (24:25): Unless the water is evaporating.

Although the values in Figure 2 to 4 should be read cautiously, considering the subjective character of identifying and scoring argumentation units, the results have a clear trend: the data gathering part of laboratory tasks does not invite argumentation discourse. Such discourse happens mainly when students try explaining observations and reflecting on the evidence for their conclusions.

(Figure 4 in about here)

What effects on argumentation relate to the strategies implemented in the tasks?*Complex data: Task 1, Container*

Considering that temperature differences identified by the groups were 1 – 3⁰C and that students sometimes disagreed over thermometer readings, there were good grounds for discussion about the evidence for deciding which container had smallest heat loss. However, all groups concluded “silver was best” regardless of how small the temperature differences were. When challenged by the teacher students did point towards “accuracy” of their data gathering. Their response reveals that “accuracy” was not interpreted in terms of *measurement error*, but rather as *measurement strategy*. They defended their conclusion with comments about measurement frequency and regularity, but never doubted that a single measurement could be wrong or uncertain. Hence, complex or “uncertain” data in itself did not create any discussion: there was no need to discuss evidence or the conclusion since measurements gave “the correct answer”.

Conflicting hypotheses: Task 2, Dissolving Salt

This strategy prompted initial discussion as intended, bringing forward conflicting views and predictions. The students, however, soon put the conceptual discussion aside and focused on data gathering. Figure 1 shows a small proportion of time spent on “hypothesising” relative to “experimentation”. Data gathering took most time and in this “mode” little attention was paid towards other parts of the investigation. Figure 1 also shows that more time in this task was spent on “co-ordination and evaluation” compared to Task 1, due to conflicts occurring when data did not give the predicted answer. Students supporting statement A (no loss of mass) expected measurements before and after dissolving the salt to be *exactly* the same, without considering differences due to measurement error, and this problem had to be solved. Several outcomes occurred, which align with Chinn and Malhotra (2002b):

Peer argumentation in the school science laboratory – exploring effects of task features

- Group A denied the problem and simply accepted their measurements without saying whether they supported or refuted the hypothesis. That is, data became the final answer to the task.
- Group B rejected their original hypothesis and concluded that salt loses mass when being dissolved in water, thus accepting data as “true”;
- Group C at first rejected their original hypothesis, but then discussed the evidence;
- Group D ignored anomalous data and kept the original hypothesis, thus accepting the hypothesis as “true” over data.

Each situation gave a different ground for argumentation. If, as in group B, students believed in the data, argumentation was short:

Ben (24:15): So we were wrong.

Bob (24:29): 183.69 [reads out the mass of the dissolved salt in solution]

Ben (24:33): Did you say it should be exactly....?

Beth (24:40): We might have missed out a bit of salt

Bob (24:55): So we might lose a bit of mass, so the hypothesis might be right

Bob (24:58): Our hypothesis was incorrect!

Beth (25:05): So it will increase a little bit, but not as much as

Beth and Bob here raise a rebuttal, but these carry little weight because data are “correct” and give the “final answer”. Group D held similar certainties about the hypothesis, ignoring the data because they *knew* their hypothesis was correct. Only Group C, who accepted data as uncertain, argued with higher frequency and advanced into more use of rebuttals. Uncertainty in the data opens up the possibility that a hypothesis might be seen as correct or not, depending on data quality.

Two conclusions arise from the conflicting hypotheses task. Firstly, presenting alternative hypotheses stimulates argumentation. This happened especially in the initial phase when students tried to resolve conflicts between hypotheses and in the experimental phase when discussing matches between hypotheses and data. Such conflicts also occurred in Task 1, where students observed the metal-surfaced container best keeping the temperature despite believing the black container was “correct”, but this task did not initiate the same amount of argumentation. We attribute the difference to explicitly-stated hypotheses in Task 2 that prompted initial discussion. Secondly, the frequency and quality of argumentation is strongly influenced by students’ understanding of uncertainty in data. Students’ belief in the “truth” of data limits the quantity and quality of argumentation.

Post-investigation discussion: Task 3

This format prompted a focused and effective discussion generating more argumentation units per unit time about the “*Container*” investigation data. The classroom, non-laboratory setting enabled students’ attention to be readily directed towards reflecting on both data and method. Surprisingly, students also paid more attention to “hypothesising”,

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3 although this was not explicitly required. Figure 1 shows almost all groups used 50% or
4 more of the time explaining *why* a metal-surfaced container should keep temperature
5 better than the white and black ones. This matter was neglected by students doing Task 1.
6 The alternative hypotheses discussion stimulated argumentation with frequent use of
7 warrants and rebuttals. Group E, for example, suggest first that the property of the metal
8 offers an explanation, but raised the rebuttal that all containers were the same type of
9 metal. They pointed out that black is known to “attract” heat. Next they suggest “layers
10 of the paint” as an explanation, but again found a rebuttal by consulting the data:
11
12

13
14 Eric (21:27): ...black might have had more paint on, but metal still does better.
15

16 Their last suggestion was that the temperature differences may be caused by “different
17 physical properties of the paint”, to which they did not manage to form any rebuttal and
18 therefore kept as their final explanation.
19

20
21 This task also showed that doing the related practical task (Task 1) first did not generate
22 more or “better” argumentation (see Table 1). In fact, some group discussions indicate
23 the opposite. Groups A to D, who did Task 1 first, compared the fictitious Task 3 data
24 with their own investigation and results, so did not see a need for discussion: their own
25 method, measuring temperature every minute, was judged the “best” strategy and their
26 own data offered the “correct” conclusion. Groups E and F, who had not done Task 1
27 beforehand, discussed the four sets of data provided, drew a conclusion and used
28 methodological criteria to decide which was experiment was best.
29
30

Discussion

31
32 Changing laboratory teaching from a “positivist” tradition (Driver et al., 2000), towards
33 nurturing authentic scientific inquiry was the motivation for this study. To date, scientific
34 argumentation, which should be a natural part of any inquiry process, has played a minor
35 role in laboratory teaching. The oversimplified methods implemented in many laboratory
36 tasks guide students from research problems to final conclusions without the need to raise
37 questions about the method used or the quality of evidence collected. Consequently, data
38 gathering becomes the main focus for student activity, while other elements of the inquiry
39 process are neglected (Abrahams & Millar, 2008; Newton et al., 1999; Watson, 2004).
40 Extant research has also shown that school experiments, for the above reason, teach
41 students a misleading picture of scientific inquiry, reinforcing an unscientific
42 epistemology by encouraging the belief that science is a simple, algorithmic form of
43 reasoning (Chinn & Malhotra, 2002a; Germann, Haskins, & Auls, 1996). Our study
44 challenges this practice by testing the effects of three different strategies in investigative
45 tasks.
46
47
48
49

50
51 The findings point to difficulties in changing existing laboratory teaching practices.
52 Despite working in a learning environment that strongly encouraged evaluation and
53 discussion of the quality of the evidence, and the task strategies, more than eighty percent
54 of students’ time and attention were focused on data gathering. Our evidence shows that
55 this “mode of working” is the least stimulating for scientific argumentation. The three
56 strategies diverted students’ attention towards “hypothesising” and “coordination and
57
58
59
60

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

evaluation” with mixed success. Students ignored complex data (Strategy 1) in Task 1 (*Container*) identifying the “best” container, without thinking of the possibility that their data may not fit this purpose. Conflicting hypotheses (Task 2, *Dissolving Salt*) produced more argumentation than Task 1, but did not for all groups. Only one group examined the evidence thoroughly, by accident, because they measured twice with different outcomes. Post-investigation discussion (Task 3) generated sufficient distance from the investigation to enable students to consider evidence in relation to the conclusion. But, even then, some students solved the task as looking for the “correct” approach and the “right” conclusion. Jiménez-Aleixandre et al. (2000) suggest that hypothetical data provided by the teacher generate different patterns of discussion than empirical, uncertain data, gathered by students themselves. A conclusion from this study is that students’ working from their own data does not *guarantee* their engaging in evaluative debate.

The study points towards two key issues in understanding why it is so difficult to engage students in scientific argumentation while working in the laboratory. First, it supports Kelly’s (2005) claim that science inquiry is an epistemic practice which requires understanding of the methodological components and criteria involved. Our findings reveal how students take data to be “true”, with no concept of “uncertainty”. We were surprised by the extent to which students put aside personal beliefs to accept a measurement uncritically. Even when admitting an error, the conditions under which the measurement was made rather than the measurement itself were responsible - a measurement was always right. Watson (2004) similarly observed:

The students carried out their tests and accepted the results of their tests as proof: the results justified claims in an unproblematic way (p 34).

This has tremendous implications for any debate regarding evaluation and coordination of evidence: if data are “true”, there is no need for argumentation. Interestingly, at times students could imply they understood that data had uncertainty, but their interpretation of “uncertainty” did not focus on measurement error, but on procedures for data gathering, such as measuring at regular intervals. On one hand, this demonstrates a failure of laboratory teaching: students learn procedures mechanistically, using them without epistemological understanding. Early laboratory teaching should focus on epistemological understanding and less on experimental procedures. Alternatively, it demonstrates that strategies trialled in this project were insufficient to change students’ understandings of the epistemic nature of scientific inquiry.

Second, of significant importance for laboratory-based scientific argumentation, is the restraint relating to “working modes”. Klahr et al.’s (1993) SDDS model offers theoretical support for this. Their “problem space” concept suggests scientists operate in three separate modes, “experimenting”, “hypothesising” and “coordination and evaluation”, where the last mode requires a combination of the first two. Our data support this model strongly. Empirically, however, we were surprised how strictly students kept to one mode, and how difficult they found it to change from one to another. Students spent most time in “experimentation” mode, which may seem natural as they are working with equipment and asked to do data gathering. “Hypothesising” could have been a

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3 natural mode, but was rarely found while students were working with equipment. Only
4 when equipment was put aside and discussions ensued, as in Task 3, was this mode
5 apparent. The reason for this, we believe, is that mental effort is required to leave one
6 mode for another. Operating in a mode means seeing the task in a particular way, being
7 aware of possibilities and restrictions for what is relevant and important. To enter another
8 mode involves changing the conception of the task, so needs stimuli. This applies when
9 going from “experimentation” to “hypothesising”, but is even more relevant for
10 “coordination and evaluation”. Students are familiar with the two first modes, but not the
11 third. Many students looked puzzled when asked to evaluate their conclusions and lacked
12 a strategy for this. Familiarising students with the mode of coordination and evaluation
13 may be a start towards establishing argumentation in the laboratory. That means
14 establishing understanding of the task *purpose* and demonstrating strategies for handling
15 the problems involved.
16
17
18
19

20 Of course, the problem of engaging students in argumentation while in the laboratory
21 could be solved partly by creating more debate in experimentation mode. Students could
22 reflect on and debate if the methods they use and data they gather are accurate, and give
23 the right information. However, for two different reasons, this can only be part of any
24 solution. Firstly, students’ investigations have to be simple. If an investigation design is
25 too complicated students lose track and their work becomes meaningless. We observed
26 examples of this during the study. Data have meaning only when students fully
27 understand the design and method from which they originate. Simple investigation
28 designs also limit how much debate could be prompted. Secondly, the argumentation we
29 aim for goes beyond ‘experimentation’. Each problem space has its own rationale for
30 argumentation. When ‘experimenting’, the problem is how to make relevant and accurate
31 observations and collect data to answer the research question. When ‘hypothesising’, the
32 problem is to find the best explanation. When doing ‘co-ordination and evaluation’ the
33 problem is to assess the certainty of the explanation/conclusion. Doing an investigation
34 answers all these three problems. If we were training laboratory technicians, the first may
35 be most important, but school students are being trained to understand disciplinary
36 critique (Ford, 2008), so the two last problems take priority.
37
38
39
40

41 Applying these perspectives to the tasks used here, and in combination with the data
42 analysis outcomes, we see potential for further development. Generally, laboratory
43 teaching should work with, not against, the problem spaces. This means guiding and
44 stimulating students’ discussions and debate towards particular questions at points in the
45 investigation where these might naturally occur. Task 1 (*Container*), for example, could
46 have included a pre-discussion similar to the ‘post-discussion’ applied in Task 3. That is,
47 students could have accessed several data sets as stimuli for discussing the best way of
48 solving the task and a likely conclusion *before* planning and conducting their own
49 investigation. This might open up rather than restrict discussion, as analysis of Task 3
50 reports. This also ensures students approach data gathering with expectations, as Task 2
51 (*Dissolving Salt*) stimulated debate. Questions asked of students while data gathering
52 should focus on technical issues only. A post-discussion added to Task 1, with an explicit
53 focus on the difference between explaining and evaluating the data would also be
54
55
56
57
58
59
60

Peer argumentation in the school science laboratory – exploring effects of task features

valuable. Hence, we can combine the best elements of all three tasks used in the study, since none alone gives the best means of stimulating argumentation in the laboratory.

Although argumentation in a laboratory context has unique features, research conducted in other contexts shows similarity. Roth and Roychoudhury (1992), De Vries, Lund and Baker (2002), Sherman and Klein (1995) and Zohar and Nemet (2002) suggest that a structured approach is important in generating productive discussion. Students need help when putting forward and identifying claims and evaluating using scientific knowledge and data. Other research suggests that establishing argumentation structures in teaching takes time, and over two weeks only limited improvement may be expected. Osborne, Erduran and Simon (2004) report their nine month intervention to be too short for students to develop sufficient skills and abilities. This is supported by Zoller, Ben-Chaim, Pentimalli and Borsese (2000), investigating first year college undergraduates' development of critical thinking. They point towards the need of recurrent opportunities to engage in same type of activities. Only when students have obtained some understanding of the pattern and purpose of activities may we expect the teaching to become efficient. The message, however, from these and other studies is that teaching of argumentation *is possible if it is explicitly addressed and taught*. The current study indicates that the same applies to the science laboratory, offering a sought-for opportunity to practice argumentation in a scientific, as opposed to the more common socio-scientific, context.

The difficulty of establishing argumentation in the laboratory, of course, should not be underestimated. Laboratory teaching is deeply rooted in the logical-empiricist tradition and has a long way to go before socio-constructivist epistemological perspectives become commonplace. If, as the current research suggests, this is hindered by psychological constraints of students operating in different "working modes", we should not be too optimistic for rapid change. More research is needed to establish better understandings of efficient strategies. The present study emphasised "implicit" teaching of epistemological criteria, and a useful alternative to be explored is what happens if these are made more explicit. A longitudinal intervention may provide revealing information. Laboratory work needs to prove less situated with "random incidents" being the main pattern. The hope is that recurrent use of task strategies may establish structures and understanding that students carry with them from one activity to the next. A last suggestion is the need for wider use of task strategies. Knowing that a teacher's ability to foster a context of argumentation is crucial (Osborne et al., 2004), the current research, with teaching being conducted by one of the researchers, has clear limitations. The strategies should therefore be tested with other teachers and across different educational cultures. This and other research, however, should be given priority as the laboratory is too important as an arena for training scientific argumentation to be left unattended.

*Peer argumentation in the school science laboratory – exploring effects of task features***References**

- Abrahams, I., & Millar, R. (2008). Does Practical Work Really Work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, 30(14), 1464-5289.
- Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honour of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum.
- Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences*, 2(2), 141-178.
- Cavagnetto, A. R. (2010). Argument to Foster Scientific Literacy: A Review of Argument Interventions in K-12 Science Contexts. *Review of Educational Research*, 80(3), 336-371.
- Chinn, C. A., & Malhotra, B. A. (2002a). Epistemologically Authentic Inquiry in Schools: A theoretical Framework for Evaluation Inquiry Tasks. *Science Education*, 86, 175-218.
- Chinn, C. A., & Malhotra, B. A. (2002b). Children's Responses to Anomalous Scientific Data: How Is Conceptual Change Impeded? *Journal of Educational Psychology*, 94(2), 327-343.
- Clark, D. B., & Sampson, V. (2008). Assessing dialogic argumentation in online environments to relate structure, grounds, and conceptual quality. *Journal of Research in Science Teaching*, 45, 293-321.
- De Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences*, 11, 63-103.
- Donnelly, J., Buchan, A., Jenkins, E., Laws, P., & Welford, G. (1996). *Investigation by Order. Policy, curriculum and science teachers' work under the Education Reform Act*. Nafferton: Studies in Education.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-312.
- Ebenezer, J., & Erickson, G. (1996). Chemistry Students' conceptions of Solubility: A Phenomenography. *Science Education*, 80(2), 181-201.
- Engel, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction* (20), 399-483.
- Erduran, S. (2008). Methodological Foundations in the Study of Argumentation in Science Classrooms. In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education. Perspectives from Classroom-Based Research*. Dordrecht: Springer
- Erduran, S., Simon, S., & Osborne, J. (2004). Tapping into Argumentation: Developments in the Application of Toulmin's Argument Pattern for Studying Science Discourse. *Science Education*, 88, 915-933.
- Ford, M. (2008). Disciplinary Authority and Accountability in Scientific Practice and Learning. *Science Education*, 92, 404 – 423.

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. *Journal of Research in Science Teaching*, 33(5), 475-499.
- Harmon, M., Smith, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V. S., Gonzalez, E. J., & Orpwood, G. (1997). Performance assessment in IEA's third international mathematics and science study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation and Educational Policy. Boston College.
- Hegarty-Hazel, E. (Ed.). (1990). *The Student Laboratory and the Science Curriculum*. London: Routledge.
- Herrenkol, L. R., Palincsar, A. S., DeWater, L. S., & Kawasaki, K. (1999). Developing scientific communities in classrooms: A sociocognitive approach. *The Journal of the Learning Sciences*, 8(3 & 4), 451-493.
- Hodson, D. (1993). Re-thinking Old Ways: Towards A More Critical Approach To Practical Work In School Science. *Studies in Science Education*, 22, 85-142.
- Hodson, D. (1996). Practical work in school science: exploring some directions for change. *International Journal of Science Education*, 18(7), 755 - 760.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: foundation for the 21st century. *Science Education*, 88, 28-54.
- Hofstein, A., Kipnis, M., & Kind, P. M. (2008). Learning in and from science laboratories: enhancing students meta-cognition and argumentation skills. In C. L. Petroselli (Ed.), *Science Education Issues and Developments* (pp. 59-94). New York: Nova Science Pub.
- Hogan, K., & Maglienti, M. (2001). Comparing epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching*, 38, 663-687.
- Jiménez-Aleixandre, M. P. (2008). Designing Argumentation Learning Environments. In S. Erduran & M. Jiménez-Aleixandre (Eds.), *Argumentation in Science Education Perspectives from Classroom-based Research*. Dordrecht: Springer.
- Jiménez-Aleixandre, M. P., Bugallo Rodriguez, A., & Duschl, R. A. (2000). "Doing the lesson" or "Doing Science": Argument in High School Genetics. *Science Education*, 84, 757-792.
- Jiménez-Aleixandre, M., & Erduran, S. (2008). Argumentation in Science Education: An Overview. In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education. Perspectives from Classroom-based Research*. Dordrecht: Springer.
- Jiménez-Aleixandre, M.P., Lopez Rodriguez, R., & Erduran, S. (2005). Argumentative quality and intellectual ecology: A case study in primary school, National Association for Research in Science Teaching, Annual Meeting. Dallas, TX.
- Katchevich, D., Mamlok-Naaman, R., & Hofstein, A. (2010). Argumentation in the Chemistry Laboratory: Inquiry and Confirmatory Experiments, NARST Annual Conference. PA: Philadelphia.
- Kelly, D. L. (2007). Discourse in Science Classrooms. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3 Kelly, G. (2005). Inquiry, Activity, and Epistemic Practice, NSF Inquiry Conference
4 Proceedings <http://www.ruf.rice.edu/~rgrandy/NSFConSched.html> . , Inquiry
5 Conference on Developing a Consensus Research Agenda Rutgers University.
6
7 Kim, H., & Song, J. (2006). The Features of Peer Argumentation in Middle School
8 Students' Scientific Inquiry. *Research in Science Education*, 36(3), 211-233.
9
10 Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A
11 developmental study. *Cognitive Psychology*, 24(1), 111-146. *Cognitive*
12 *Psychology*, 24(1), 111-146.
13
14 Kolstø, S. D., & Mestad, I. (2005). Learning about the nature of scientific knowledge: the
15 imitating-science project. In K. Boersma, M. Goedhart, O. D. Jong & H. Eijkelhof
16 (Eds.). *Research and the Quality of Science Education*, 247-258. Dordrecht:
17 Springer.
18
19 Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*.
20 Cambridge, MA: MIT Press.
21
22 Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, 62, 155-178.
23
24 Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: Univ. of Chicago
25 Press.
26
27 Lemke, J. (1990). *Talking Science: Language, Learning, and Values*. Norwood, NJ:
28 Ablex.
29
30 Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Learning and Teaching in the
31 School Science Laboratory: An Analysis of Research, Theory, and Practice. In S.
32 K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education*.
33 Mahwah, NJ: Lawrence Erlbaum Associates.
34
35 Martin, A., & Hand, B. (2009). Factors affecting the implementation of argument in the
36 elementary science classroom. A longitudinal study. *Research in Science*
37 *Education*, 39, 17-38.
38
39 Millar, R., Lubben, F., Gott, R., & Duggan, S. (1994). Investigating in the school science
40 laboratory: conceptual and procedural knowledge and their influence on
41 performance. *Research Paper in Education*, 9(2), 207 — 248.
42
43 Mork, S. M. (2005). Argumentation in science lessons> Focusing on the teacher's role.
44 *Nordic Studies in Science Education*, 1(17-30).
45
46 Murphy, P., & Gott, R. (1984). *The Assessment Framework for Science at Age 13 and*
47 *15*. APU Science report for teachers: 2. London: Department for Education and
48 Science.
49
50 Naylor, S., Keogh, B., & Downing, B. (2007). Argumentation and Primary Science.
51 *Research in Science Education*, 37, 17-39.
52
53 Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the
54 pedagogy of school science. *International Journal of Science Education*, 21(5),
55 553-576.
56
57 Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in
58 school science. *Journal of Research in Science Teaching*, 41(10), 994-1020.
59
60 Qualifications and Curriculum Authority. (2007). *Science. Programme of study for key*
stage 3 and attainment targets. The National Curriculum 2007 (pp. 206-219).
London: QCA.

Peer argumentation in the school science laboratory – exploring effects of task features

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Richmond, G., & Striley, J. (1996). Making meaning in classroom. Social processes in small group discourse and scientific knowledge building. *Journal of Research in Science Teaching*(33), 839-858.
- Roth, W. -M., & Roychoudhury, A. (1992). The social construction of scientific concepts or the concept map as conscription device and tool for social thinking in high school science. *Science and Education*, 76, 531–557.
- Sadler, T. D. (2004). Informal Reasoning Regarding Socioscientific Issues: A Critical Review of Research. *Journal of Research in Science Teaching*, 41(5), 513-536.
- Sadler, T.D., Chambers, F.W. & Zeidler, D.L. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education*, 26(4), 387-409.
- Sampson, V., & Clark, D. B. (2008). Assessment of Ways Students Generate Arguments in Science Education: Current Perspectives and Recommendations for Future Directions. *Science Education*, 92(3), 447-472.
- Sandoval, W. A., & Millwood, K. A. (2007). What can argumentations tell us about epistemology? In S. Erduran & M. Jimenez-Aleixandre (Eds.), *Argumentation in Science Education*. Dordrecht: Springer.
- Sandoval, W. A., & Reiser, B. J. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3), 345-372.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. New York: Erlbaum.
- Schofield, B., Black, P., Head, J. & Murphy, P. (1985) *Science in Schools: Age 13 Research Report No. 2* London: Department for Education and Science
- Séré, M.-G., Leach, J., Niedderer, H., Psillos, D., Tiberghien, A., & Vicentini, M. (1998). *Improving Science Education: issues and research on innovative empirical and computer-based approaches to labwork in Europe*. Final report from Labwork in Science Education. .
- Sherman, G. P., & Klein, J. D. (1995). The effects of cued interaction and ability grouping during cooperative computer-based science instruction. *Educational Technology Research and Development*, 43, 5–24.
- Simon, S., Erduran, S., & Osborne, J. (2006). Learning to Teach Argumentation: Research and development in the science classroom. *International Journal of Science Education*, 28(2-3), 235-260.
- Simonneaux, L. (2001). Role-play or debate to promote students' argumentation and justification on an issue in animal transgenesis. *International Journal of Educational Research*, 23, 902-927.
- Tapper, J. (1999). Topics and manner of talk in undergraduate practical laboratories *International Journal of Science Education*, 21(4), 447 – 464
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Watson, J. R., Swain, J. R. L., & McRobbie, C. (2004). Students' discussion in practical scientific inquiries. *International Journal of Science Education*, 26(1), 24-45.
- Welford, G., Bell, J., Davey, A., Gamble, R. & Gott, R. (1986) *Science in Schools Age 15: Research Report No. 4* London: Department for Education and Science

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3 Yerrick, R. K. (2000). Lower track science students' argumentation and open inquiry
4 instruction. *Journal of Research in Science Teaching*, 37, 807-838.

5
6 Zeidler, D., Sadler, T., Simmons, M., & Howes, E. (2005). Beyond STS: A research-
7 based framework for socioscientific issues education. *Science Education*, 89,
8 357–377.

9
10 Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumenation skills
11 through dilemmas in human genetics. *Journal of Research in Science Teaching*,
12 39, 35-62.

13 Zoller, U., Ben-Chaim, D., Pentimalli, R., & Borsese, A. (2000). The disposition towards
14 critical thinking of high school and university students: An inter-intra Israel-
15 Italian study. *International Journal of Educational Research*, 22, 571-582.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

*Peer argumentation in the school science laboratory – exploring effects of task features***Appendix 1 Task 1, Container****Container**

A company transporting hot liquid is buying a new tanker lorry and wants to know what colour the tank should be. You are asked to do an experiment to help give them advice if the colour has any effect on loss of heat.

You should find out *if* the colour has any effect and, if so, *what colour* (black, white or metal) best keeps the temperature.



The company wants to be sure about its decision and therefore would like each scientist to tell what he or she thinks about the experiment. You should:

1. Do the experiment to get the data.
2. Discuss how big the differences are between the containers and how sure you can be that the results are correct.
3. Make an advice to the company, telling them if you all agreed or if there were different opinions.

Your answer to the company must be supported by data and you should try to explain the results.

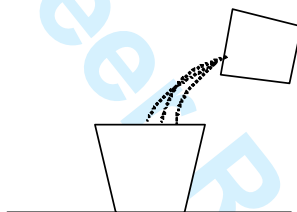
You will have:

- Three types of containers
- Hot water
- Thermometers
- A timer

Appendix 2 Task 2, Dissolving Salt**Dissolving salt**

Some students are investigating what happens to the mass of a container with water when we dissolve salt in it. They have different explanations. With whom do you agree? You should:

1. Discuss each statement and tell **if and why** you think it is right or wrong.
2. Select two statements you are uncertain about or you think are wrong.
3. Then do the experiments to test these two statements.
4. Discuss if the experiments are good enough to draw a conclusion and, if so, why or why not.



Statements:

- a) *When salt dissolves it still weighs the same. The total mass is therefore exactly the same as the water and the salt added.*
- b) *There is always a small loss of mass when something dissolves. The mass will increase but less than the mass of the salt added.*
- c) *The higher the temperature the more salt dissolves, the new mass therefore depends on the temperature of the water.*

You will have:

- Containers
- Hot and cold water
- Thermometers
- A stirrer
- A scale

*Peer argumentation in the school science laboratory – exploring effects of task features***Appendix 3 Task 3 Post-investigation discussion of Container task****Container**

Some students have been doing an experiment to find out if the colour of an object has any effect and on its heat loss, and if so, what colour best keeps the temperature.

They had containers in three different colours: black, white and metal (not painted). In their experiments they filled the containers with hot water and measured the temperature as the containers cooled down. The tables show results from four different groups. They all did the experiment in different ways.

Group A			
150 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
1	90	90	91
2	84	84	85
3	80	79	80
4	76	75	76
5	72	72	73
6	71	70	72
7	70	69	71

Group B			
100 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
5	75	76	78
10	55	57	48
15	40	42	44
20	30	34	36
25	30	27	30
30	23	24	26
35	19	18	19

Group C			
150 ml water in each container			
Time (min)	White	Black	Metal
0	96	97	97
1	90	90	91
5	72	72	73
16	40	42	44
21	30	31	35

Group D			
100 ml water in each container			
Time (min)	White	Black	Metal
5	75	77	77
10	55	57	58
15	40	42	44
20	29	33	35
30	21	20	22

1. Decide which group you think did the best experiment. Explain why.

Continues

Peer argumentation in the school science laboratory – exploring effects of task features

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2. Analyse the results from the group you think did the best experiment and decide which container has least heat loss (best keeps the temperature).

My conclusion is the

Black White Metal coloured They are the same

3. How certain are you about the conclusion? Tick one of the boxes below, or make your own statement.

- I am absolutely certain that the container I identified is the right one.
- The container I have identified is the best in the data, but the group's data gathering is too poor to make a final conclusion.
- The container I have identified is the best in the data, but the results are very uncertain. Another test is likely to show different results.
- It is impossible from the data to conclude if any container best keeps the temperature.
- Other: _____
- _____
- _____

4. Compare your answers to the other students in your group. Decide between you what answers you think are correct.

Our final conclusion is that the _____ container(s)
best keep(s) the temperature, because

*Peer argumentation in the school science laboratory – exploring effects of task features***Tables and figures:**

	<i>Task 1: Container</i>	<i>Task 2: Dissolving</i>	<i>Task 3 Discussion of Container task</i>
Group A (Girls)	x	x	x
Group B (Mixed)	x	x	x
Group C (Mixed)	x	x	x
Group D (Boys)	x	x	x
Group E (Mixed)			x
Group F (Boys)			x

Table 1: Design of the study

Level 1	Level 1 argumentation consists of arguments that are a simple claim versus a counter-claim or a claim versus a claim.
Level 2	Level 2 argumentation has arguments consisting of a claim versus a claim with either data, warrants, or backing but do not contain any rebuttals.
Level 3	Level 3 argumentation has arguments with a series of claims or counter-claims with either data, warrants, or backing with the occasional weak rebuttal.
Level 4	Level 4 argumentation shows arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counter-claims.
Level 5	Level 5 argumentation displays an extended argument with more than one rebuttal.

Table 2: Erduran et al.'s (2004: 928) Analytical Framework for Assessing Quality of Argumentation

Peer argumentation in the school science laboratory – exploring effects of task features

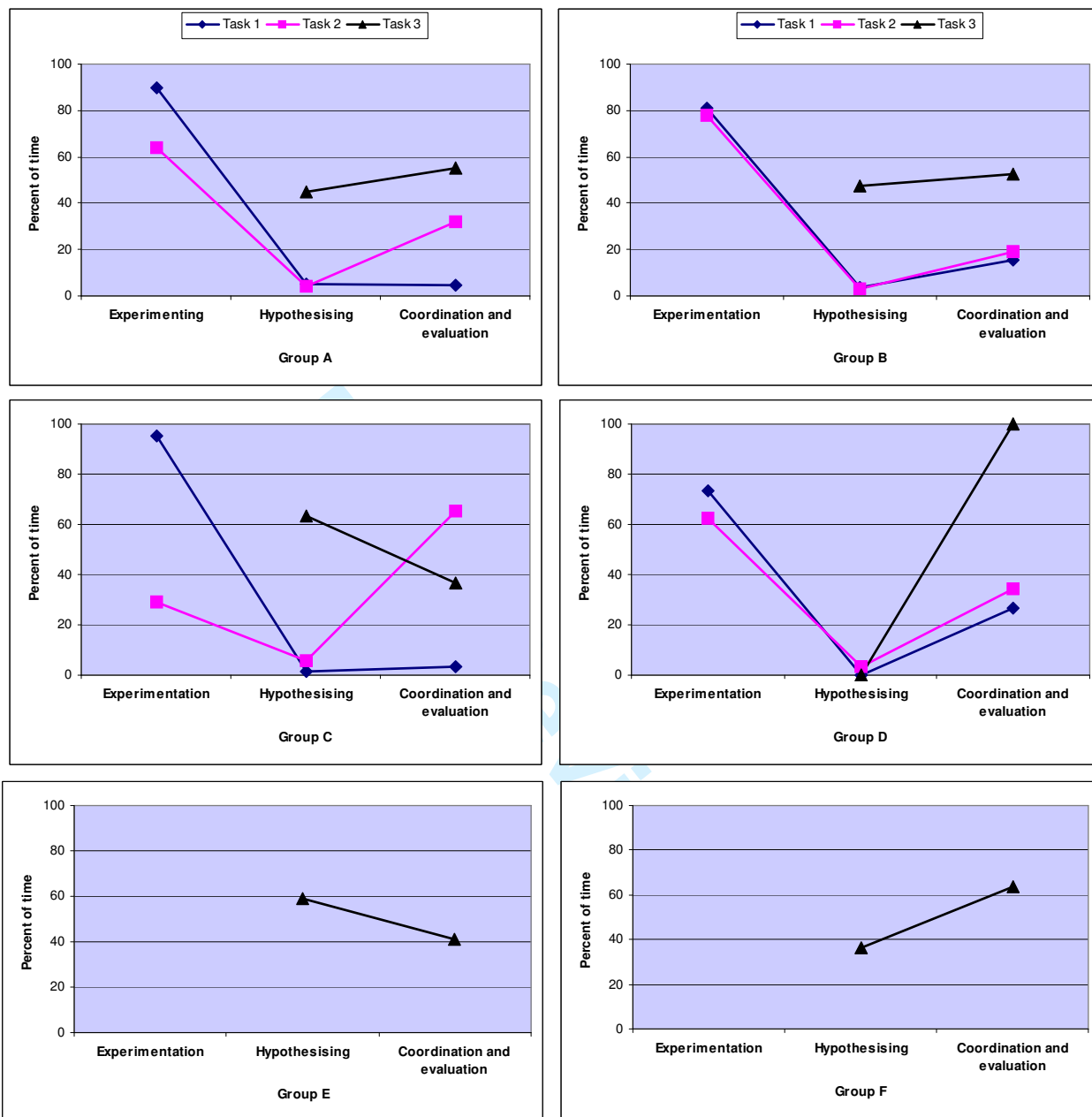


Figure 1: Orientation data: Total time spent by six student groups A-F on experimenting, hypothesising and co-ordinating and evaluating over three different tasks

Peer argumentation in the school science laboratory – exploring effects of task features

	<i>Task 1</i>	<i>Task 2</i>	<i>Task 3</i>	Total
Group A	7	10	8	25
Group B	4	7	6	17
Group C	9	17	3	29
Group D	3	6	1	10
Group E			11	11
Group F			2	2
Sum Gr A to D	23	40	18	81

Table 3: Number of argumentation units identified in each task each group

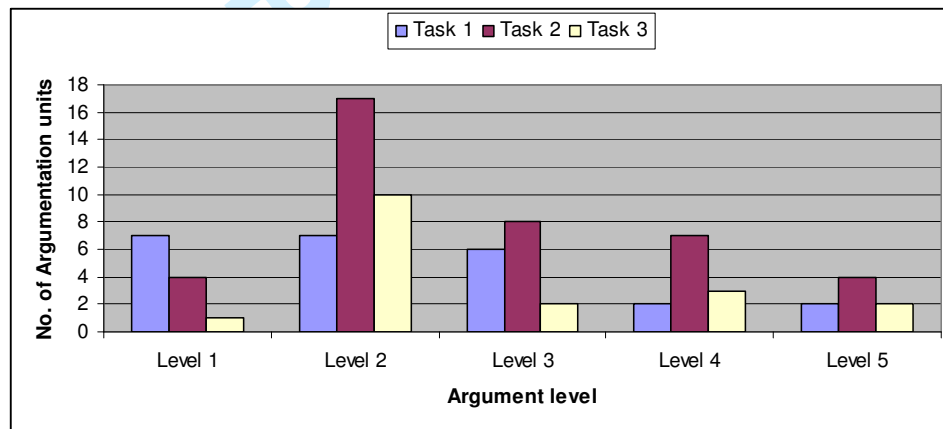


Figure 2: Number and level of argumentation units per task for Groups A to D.

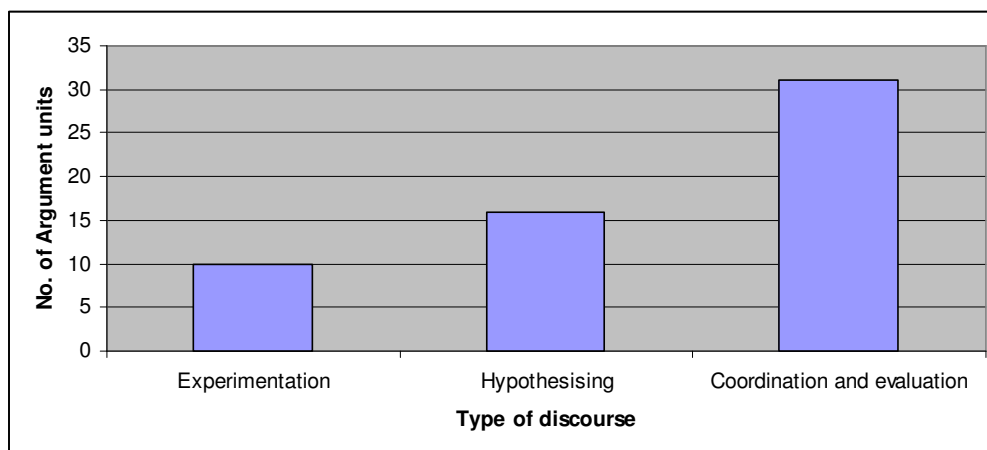
Peer argumentation in the school science laboratory – exploring effects of task features

Figure 3: Number of argumentation units summarised within each type of discourse for groups A to D.

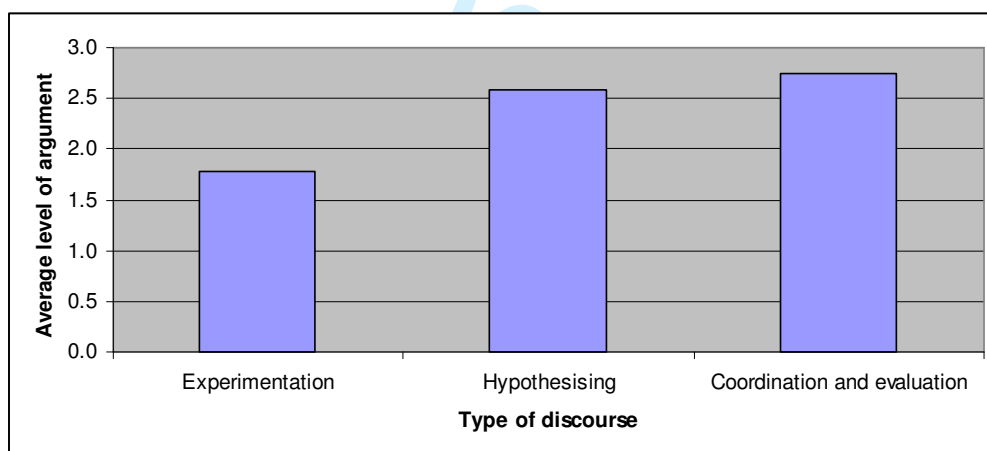


Figure 4: Average level of argumentation within each type of discourse for groups A to D.