International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: https://www.tandfonline.com/loi/tsed20

Gauging Students' Untutored Ability in Argumentation about Experimental Data: A South African case study

Fred Lubben , Melanie Sadeck , Zena Scholtz & Martin Braund

To cite this article: Fred Lubben , Melanie Sadeck , Zena Scholtz & Martin Braund (2010) Gauging Students' Untutored Ability in Argumentation about Experimental Data: A South African case study, International Journal of Science Education, 32:16, 2143-2166, DOI: 10.1080/09500690903331886

To link to this article: https://doi.org/10.1080/09500690903331886

Published online: 16 Nov 2009.

International Journal of Science Education
Vol. 32, No. 16, 1 November 2010, pp. 2143–2166

RESEARCH REPORT

Gauging Students' Untutored Ability in Argumentation about Experimental Data: A South African case study

Fred Lubben*, Melanie Sadeck, Zena Scholtz and Martin Braund
Cape Peninsula University of Technology, South Africa

This paper reports on a study into the untutored ability of Grade 10 students to engage in argumentation about the interpretation of experimental data, and the effect of unsupported small group discussions on this ability. In this study, argumentation intends to promote critical thinking. The sample includes 266 students from five South African schools across the resource spectrum, forming 70 friendship groups. Students are required to provide written interpretations of experimental data, and justify these interpretations based on the evidence and concepts of measurement. Individual responses form the basis of small group discussions after which students again provide written justified interpretations of the readings. The data show an initial low level of argumentation, with significant variation according to school resource levels. Considerable improvement in the level of argumentation occurs merely through small group discussions unsupported by the teacher. The findings suggest several factors influencing argumentation ability such as experience with practical work, perceptions of the purpose of small group discussions, the language ability to articulate ideas and cultural influences. Methodological issues arising from the study and implications for teaching and assessment are discussed.

Keywords: *Argumentation; Experimental evidence; Critical thinking; Science resource level; Secondary education; Small group discussions; South Africa*

Introduction

This paper reports part of a larger study on the introduction of critical thinking in South African schools. In particular, this report focuses on argumentation as part of critical thinking, and explores the learning effect of a specific strategy often used in the teaching of argumentation, that is small group discussions, independent from scaffolding teaching interventions aiming specifically at strengthening argumentation

*Corresponding author. Department of Educational Studies, University of York, Heslington, York YO1 5DD, UK. Email: fel1@york.ac.uk

ISSN 0950-0693 (print)/ISSN 1464-5289 (online)/10/162143–24

DOI: 10.1080/09500690903331886

skills. It is intended that the findings will contribute to the development of suitable teaching interventions to promote argumentation abilities building on the improvement already achieved through small group discussions *per se*.

This study draws on two—partly overlapping—bodies of research, that is the effectiveness of teaching methods for the improvement of students' argumentation abilities, and the learning effect of small group discussions. We will start with summarising our view of the position of argumentation within the area of critical thinking.

The Relationship between Critical Thinking and Argumentation

When a new curriculum was introduced in Grades 10–12 in South African schools in 2006, considerable emphasis was laid on developing students' abilities as critical and creative thinkers (DoE, 2004). The South African Department of Education (DoE, 2004), like curriculum development agencies in other parts of the world, stipulates that the school science curriculum involves "critical inquiry, reflection, and the understanding of [science] concepts and processes, and their application in society" (p. 46). If teaching is purported to be constructivist, in the sense of building on students' existing ideas (Driver, Asoko, Leach, Mortimer, & Scott, 1994), it is crucial to document the ways students cope with situations where critical thinking is required.

Critical thinking may be viewed as a set of cognitive skills used for decision making in complex situations (Paul & Elder, 2004) or for solving "ill-structured" problems (Kuhn, 1999). Critical thinking has often been described in relation to creative thinking. Harris (1998) contrasts these two thinking modes: critical thinking as "vertical", linear, objective, and convergent thinking aimed at strengthening a hypothesis or claim, and creative thinking as "horizontal", lateral, associative, subjective, and divergent thinking aimed at mapping out the terrain around an assertion or a claim. In contrast, Paul (1990) emphasises that critical thinking requires imagination of potential arguments or possible evidence to strengthen or counter an assertion or a claim. He stresses that creative thinking is, thus, an essential component of critical thinking.

With respect to the debate as to whether critical thinking should be seen as subjective and/or objective, Kuhn (1999) draws attention to different levels of meta-knowing about the nature of assertions. At a basic "realist" level, assertions are seen as copies of an external reality, which is directly knowable since knowledge is certain and objective—in this case critical thinking is unnecessary. At an "absolutist" level, assertions are facts (correct or incorrect) representing reality, which (again) is knowable based on certain and objective knowledge—in this case critical thinking determines truth or falsehood. At the "multiplist" level, assertions are seen epistemologically as opinions only accountable to their owners, with the premise that knowledge is subjective and uncertain—here critical thinking is irrelevant. Only at the "evaluative" level of epistemological understanding, where assertions are considered to be judgements using criteria of argument and evidence, and where knowledge is seen as subjective and uncertain, critical thinking is valued as a meta-cognitive tool to strengthen assertions and enhance domain-specific understanding. This particular understanding of the nature of assertions as judgements coincides with essential components of the nature of science (see, for instance, Abd-El-Khalick, Lederman, Bell, & Schwartz, 2002; Tsai & Liu, 2005) and thus makes the development of critical thinking relevant to science education.

But what is included in critical thinking? Paul's (1990) integrationist view of critical thinking with creative thinking also emerges from the Delphi study of expert views of the content of critical thinking (Facione & Facione, 1992) which lists the following six key critical thinking skills, each with a number of sub-skills:

interpretation of information (clarifying meaning, categorisation, decoding significance);
analysis of evidence (examining ideas, identifying and analysing arguments);
evaluation (assessing claims and arguments);
inference (querying evidence, conjecturing alternatives and drawing conclusions);
explanation (stating results, justifying procedures, presenting arguments);
self-regulation (self-examination and self-correction).

Argumentation is seen as a major way of operationalising critical thinking in specific knowledge domains such as science (Kuhn, 1993). Another operationalisation could be statistical reasoning as reported by Osana and Seymour (2004). The process of argumentation centres around the generation and evaluation of evidence to support an assertion. Thus, the process of argumentation draws heavily on the critical thinking skills (2)–(4) listed above, that is analysing and evaluating evidence and drawing inferences. The process of argumentation includes the abilities to formulate an argument, using evidence to support the argument and, even stronger, to counter arguments of others. Argumentation thus requires linear reasoning (justifying the selection of evidence to support a claim) as in Confucian thinking, and also back-and-forward dialogic thinking (Andrews, 2007). In Harris' (1998) terms, argumentation requires both critical and creative thinking. In a study of students' antilogos ability (the ability to evaluate whether warrants may support a variety of claims), Glassner and Schwarz (2007) find that critical and creative thinking skills are interdependent. They conclude that earlier findings of independence of these two types of thinking are related to the (non-argumentative) tasks used to test the relationships, and that argumentation forms the bridge between creative and critical thinking.

In science education considerable work has been done to survey the ideas of students at different ages about the relationship between data (or other types of evidence) and claims. Leach, Driver, Millar, and Scott (1997) report that school students as old as 16 years of age do not differentiate between observations (the data) and conclusions (the theory). Only at undergraduate level (Dagher, Brickhouse, Shipman, & Letts, 2004; Ryder, Leach, & Driver, 1999), students explicitly use data to identify patterns and draw conclusions.

Dealing with evidence in a critical way should therefore be central in the teaching of school science (Kuhn, 1993) and could be argued as being a more intellectually honest representation of science as it recognises the ways in which scientific knowledge has been constructed and validated (Duschl & Osborne, 2002; Sandoval & Reiser, 2004).

Promoting Argumentation Abilities

In their overview article, Jiménez-Aleixandre and Erduran (2008) propose five intertwined contributions which may emerge from the introduction of argumentation in the science classroom (p. 5). These include:

1. supporting access to cognitive and meta-cognitive processes such as modelling expert performance;
2. achieving scientific literacy which will assist students to talk or write the language of science;
3. supporting the enculturation into the practices of science especially scientific knowledge evaluation;
4. encourage the development of reasoning, particularly the choice of theories based on rational criteria;
5. strengthening communicative competences, in particular critical thinking.

The authors acknowledge that these contributions influence one another, yet they are useful categories in situating our own work on argumentation as a tool for developing critical thinking amongst the last two of these contributions.

In discussing different frameworks for describing the nature and quality of student-generated arguments, Sampson and Clark (2008) differentiate between domain-general frameworks, such as those developed by Toulmin (2003) and Schwarz, Neuman, Gil, and Ilya (2003), and domain-specific analytical frameworks (for instance, Kelly & Takao, 2002; Lawson, 2003; Sandoval & Millwood, 2005; Zohar & Nemet, 2002). Since critical thinking, and thus argumentation, features across different subject curriculum descriptions in South Africa, a domain-general analysis framework was adopted for this study, in particular the scheme developed by Toulmin (2003).

In Toulmin's model, arguments are interrogated for the presence of six different components. Firstly, claims are general propositions or assertions. The term data refers to experimental data or other evidence from sources such as the internet, library or the media. Warrants are reasons used to support a given claim, whereas backings are (scientific) theories or other generalisations such as moral codes used to support a warrant. A rebuttal is a reasoned objection to a warrant or a backing to challenge their validity. A qualifier delineates the generalisability of the claim, indicating the degree of certainty that can be attributed to it. One of the persistent criticisms of the Toulmin model (see, for instance, Andrews, 2007; Duschl, 2008; Jiménez-Aleixandre, Rodríguez, & Duschl, 2000) has been the lack of distinctiveness between warrants, backings and sometimes qualifiers with recurrent need for detailed exemplification of how researchers/teachers have interpreted the different components for students' arguments.

Nevertheless, Toulmin's model has been used by several science educators for the analysis of the nature and quality of students' arguments in response to carefully designed argumentation tasks (for instance, Jiménez-Aleixandre et al., 2000; Kelly & Chen, 1999; Osborne, Erduran, & Simon, 2004). However, none of these studies used Toulmin's scheme for tasks requiring arguments for interpreting experimental data.

The teaching of argumentation, normally, makes extensive use of small group discussions. For instance, Osborne et al. (2004) have developed learning materials for argumentation tasks for such discussions, and the same researchers developed teaching strategies for supporting such small group discussions (Erduran, Osborne, & Simon, 2004). This paper, however, explores the effect of small group interaction on argumentation abilities without explicit student preparation for argumentative discourse, and without teaching strategies to specifically develop the level of argumentation or collaborative discourse skills. Thus, finally we turn to summarising some of the body of knowledge related to the use of evidence in small group discussions.

The Use of Evidence in Small Group Discussions

Research on the learning outcomes of small group discussions in science education has concentrated on changes in participants' group collaboration skills (see, for instance, Mortimer & Scott, 2003; Richmond & Striley, 1996), their understanding of the nature of science (see, for instance, Tao, 2003) and the development of conceptual understanding (Mercer, Dawes, Wegerif, & Sams, 2004). Most relevantly, some of the latter research explores specifically the cognitive effects of small-group work in the area of the "understanding of evidence" (Chan, 2001; De Vries, Lund, & Baker, 2002; Hogan, 1999). For the purposes of this paper, *understanding of evidence* is defined as the understanding "related to the collection, validation, representation and interpretation of evidence" (Gott & Duggan, 1996, p. 793), that is the ability to co-ordinate (primary or secondary) data to construct theory (models or new concepts).

Finkel (1996) and Chan (2001) dealt with understanding of evidence by exploring the effect of the introduction of external conflict within the discussion task through the provision of incompatible information. Both studies conclude that such a stimulus with a built-in conflict intensified the group interaction, and improved participants' understanding of evidence, particularly their use of evidence to reconstruct conceptual models. However, both studies suggest that incompatible task information needs to be supplemented by mechanisms of foregrounding incompatible reasoning of group members. Sherman and Klein (1995), Lajoie, Lavigne, Guerrera, & Munsie (2001) and De Vries et al. (2002) report on the effect of group discussions supported by educational software packages. They provide evidence that specific features of these packages such as explicit discussion cues in the learning sequence (Sherman & Klein, 1995) or on-line tutor feedback pointing out differences in reasoning of group members (De Vries et al., 2002) improved the cognitive effect of group interactions. Tolmie and Howe (1993) draw attention to the differences in managing incompatible information within differently gendered groups. They conclude that in all-female groups the sense-making efforts focused on finding common ground between the reasoning suggested by individual group members. In contrast, in all-male groups the emphasis was on defining differences between the reasoning of peers and determining the merit of each.

More recently, Sampson and Clark (2009) examined the impact of collaboration during scientific argument. Their findings suggest that groups of students did not produce better arguments than students who worked alone, yet a substantial proportion of students adopted at least some elements of the group's argument and that collaborative conditions demonstrated superior performance on mastery and transfer problems.

Hogan (1999) looked at the meta-cognitive effect of small group discussions. She reports that the use of such discussions supported by a specific teaching programme fostering collaborative reasoning (including the reasoning required to evaluate and strengthen knowledge claims) improved students' meta-cognitive knowledge, including their knowledge of reasoning about evidence. This improvement in collaborative reasoning depended on students' perspective on learning: students with a *learner-as-explorer* perspective gained significantly more than peers with a *learner-as-student* perspective (Hogan, 1999). The studies mentioned above report on interventions that combine the use of group discussions with additional teaching designs, such as the creation of an external conflict, computer assistance, or training in collaborative reasoning. Little is known about the effect of group discussions *per se* on students' understanding of evidence. To what extent would the same effects have been achieved without the additional teaching designs? The current study takes one step back and explores the effect of unsupported small group discussions on students' ability to use evidence.

A second aspect of the effect of small group work in science learning has been under-reported, that is the influence of culture on small group discussions by students in African contexts. There is a likelihood that cultural perspectives may influence the outcome of small group discussions in science classes. For instance, in an evaluation of a learner-centred intervention in Botswana, Tabulawa (1997) concludes that the emphasis on generating knowledge through group discourse conflicts with the cultural perception of the role of authority in knowledge construction. This influence is evident in the continuous request for confirmation from the external authority, that is the teacher or the book (see also Koosimile, 2004). The influence of authority equally emerges in the perception that the weight of a discussion statement is closely related to the status of the group member making such a statement. Similarly, there is some evidence that the aspiration towards consensus and inclusivity (Nsamenang, 1999; Venter, 2004) may influence the conventions and outcomes of group discourse (see, for instance, Scholtz, Braund, Hodges, Koopman, & Lubben, 2008).

This paper explores the contribution of small group discussions to the improvement of individual student's ability in argumentation. The study addresses the following research questions:

- What is the untutored ability in argumentation of students of Grade 10 (usually varying from 15 to 20 years of age) in South African schools?
- To what extent do small group discussions help Grade 10 students in South African schools in developing their ability in argumentation?
- What factors seem to hinder or encourage the ability in argumentation in these circumstances?

Research Methods

The sample consisted of 266 Grade 10 students, organised in 70 friendship groups of normally four students each. These groups were drawn from eight classes in five schools across the Cape Town metropolitan area. Schools were identified for reasons of convenience, such as having been involved in previous curriculum projects. The legacy of "apartheid" with its hierarchical structures and disparate socio-economic contexts (Steyn, 2000) is deep-rooted and lingers on, despite attempts by the National Education Department to implement an equal education system in South African schools. In practice, real differences in schools' resource levels prevail (Soudien, 2007). Within this context, three of the five schools were recognised as previously advantaged. They were better resourced with qualified and experienced science staff, well-equipped laboratories and a range of books and teaching aids, a structured curriculum implementation and students whose first language coincides with the teaching language, that is English or Afrikaans. Students came from established middle-class areas. Even though the background of these students may have been mixed in terms of their religion or race, the cultural practices were largely Western European. The two remaining schools were recognised as previously disadvantaged. They were both situated in a newly expanded urban area where the accommodation consisted of tin shanties, erected by the people themselves, and small brick houses. Many of the parents or grandparents would have migrated from the rural parts of South Africa in search of a better life in the urban area. Xhosa cultural influences and practices are maintained, for example respect for the elders is considered very important. These two schools, situated in a poor socio-economic environment, were less resourced having qualified but less experienced science teachers; limited access to textbooks, internet and laboratories; and large-size classes of students whose mother tongue (isiXhosa) differs from the medium of instruction (English).

Since we were interested in argumentation abilities of individual students, we collected written rather than oral data, as the former are less dependent on contributions of peers' in-group interactions. The argumentation task (see Appendix) consisted of two given sets of experimental readings for repeated measurements of the distance a ball will roll off a slope. Each set of readings was provided with their (slightly different) average values. Cartoon figures in the task sheet posed three opposing claims: that the data sets were similar, that they were not similar and that there is a third view. The content validity of the task had been tested with slightly older students (Lubben, Campbell, Buffler, & Allie, 2001).

First, individual group members were asked to write down which of the claims they agreed with, and to support their choice. These written responses were read out to the group and they then formed the basis for a subsequent group discussion. The language of the group discussion was not prescribed, but most groups in the better-resourced schools used English or Afrikaans (also the teaching language), whereas almost all groups in the less-resourced schools used IsiXhosa with frequent code-switching into English (the teaching language) for terminology and occasionally for expressions related to measurement. Afterwards, individual group members again wrote their views about the similarity of the experimental data sets, justifying the reasons for their choice.

The instrument was refined and validated through an initial pilot with 93 Grade 12 students of similar diversity of backgrounds. Explicit guidelines were developed to guide teachers when doing the activity with their students. This ensured uniformity in instructions given to learners as well as in the sequence in which the tasks were completed. The teachers were asked not to do the practical beforehand, as had been the case for one of the schools in the pilot. On the basis of the pilot the language of the task was simplified. The low frequencies of higher levels of argumentation in the pilot were of concern. Thus, a "writing frame" (Wray & Lewis, 1997) was added requesting students give reasons why they do not support the alternative views provided in the task. This allowed students an opportunity to document rebuttals, and so to increase the explication of higher levels of argumentation (if available).

The pilot responses were also used to confirm the suitability of a coding scheme using levels of sophistication of argumentation developed by Erduran et al. (2004). Their coding scheme is based on Toulmin's (2003) model of argumentation and identifies the extent to which arguments are developed, claims are substantiated and claims made by others are challenged. These levels of argumentation are summarised in Table 1.

These levels of argumentation were used for double-blind coding by two researchers of 25% of the responses in the sample. An 81% inter-coder agreement indicated a high coding reliability. Frequencies of initial levels of argumentation were used to

Table 1. Levels of argumentation (from Erduran et al., 2004)

Levels	Description
1	Arguments that are a simple claim versus a counter-claim or a claim versus a claim
2	Arguments consisting of a claim versus a claim with either data, warrants or backings but no rebuttals
3	Arguments with a series of claims or counter-claims with either data, warrants or backings with the occasional weak rebuttal
4	Arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counterclaims
5	Displays an extended argument with claims supported by data and warrants, with more than one rebuttal

survey students' untutored ability of argumentation before they engaged in group discussions. Frequencies of levels of argumentation and the nature of changes in these levels after the small group discussions indicate students' ability to incorporate evidence to strengthen an argument. In cases where obvious, but tacit, claims are made these have been inferred. For instance, several respondents noted the differences in the averages of the experimental results derived by the two groups in the task, but their response only focused on suggesting how to decide whose result was correct. Such responses have been coded as making the tacit claim that the results of both groups did not agree. Also, where one summary statement was provided for the post-discussion views of a whole group, these views have been attributed to each member in the group. The validity and implications of this analysis strategy will be considered more extensively in the discussion section below on methodological issues.

Results

Students' Use of Levels of Argumentation

Table 2 provides a summary of the ability in argumentation of Grade 10 students before engaging in group discussion, separated by the resource level of their schools.

The data in the last column of Table 2 show that more than half of the students (145 out of 266, thus 55%) provide an argument at Level 2, whereas almost one in three students (29%) make a claim with no supporting evidence, typical of Level 1. A very small proportion of students spontaneously provide rebuttals, characteristic of the higher levels. The table also suggests that the levels of argumentation differ for students in less-resourced and better-resourced schools. Whereas 85% of the students in better-resourced schools use Levels 2 and above, this percentage is about half (43%) for those in less-resourced schools. The few instances of the use of Levels 3 and 4 only occur in better-resourced schools. Using Mann-Whitney statistics, these data show that the distribution in the two groups differs significantly (Mann-Whitney $U = 8422.0$, $n_1 = 142$, $n_2 = 87$, $p < 0.001$ two tailed). The difference could

Table 2. Pre-discussion levels of argumentation of students in less-resourced and better-resourced schools

	Frequencies of students (%)		
Levels of argumentation	In less-resourced schools (n = 178)	In better-resourced schools (n = 88)	Total (n = 266)
Level 1	65 (37)	12 (14)	77 (29)
Level 2	77 (43)	68 (77)	145 (55)
Level 3	—	3 (3)	3 (1)
Level 4	—	4 (5)	4 (1)
Level 5	—	—	—
Unclassifiable	36 (20)	1 (1)	37 (14)

be explained by the much larger proportion of unclassifiable responses from students at less-resourced schools (20% against 1%).These data may suggest a handicap in the use of English as the language of learning. It seems that coherent formulation of complex arguments may require a level of written language sophistication more prevalent in the better-resourced schools. The quotations in the following sections intend to illustrate the different argumentation levels of student responses, and at the same time they amplify the influence of written language proficiency.

Levels of Argumentation

The following paragraphs illustrate the students' argumentation ability by presenting quotations of their written arguments representative of the different levels 1–4 in Erduran et al.'s (2004) scheme. Since the quotes illustrate the argumentation structure and not the substance of the argument, they are drawn from across the choices A–C, that is the claims, made by the students. Each quote is presented *verbatim* and followed by a student's identification, with the first number indicating the school, the second the group and the subsequent letter identifying the group member (a–d). As contextual information, schools 1–3 are classified as better resourced, and schools 4–5 are less so.

Level 1 argumentation. Level 1 argumentation emerged in two main ways. Firstly, this level was allocated when students repeated one of the claims they were invited to select without much supporting amplification. Typical examples are as follows:

> I choose B, because A is not the same with B, and B it is the correct answer. (4/12b)

> I think [it] is group C because they are not agree with any of this average. They just have another view which they will explain. (3/24b)

> We choose C because there will be no solution if these two groups do not agree. That is why I choose C, because C has another view to explain these two groups to be combined. (4/17b)

Secondly, a much larger cluster of students focused on identifying the correct average. This concern had its origin in the claim that the averages provided did not agree or were not the same. Several students stated this claim explicitly before suggesting ways of identifying the "correct" average. Typical arguments are the first two quotes provided below. In addition, several arguments used an implicit claim that the averages for Groups A and B did not agree as a reason for the need to determine the "correct" average, as illustrated by the subsequent quotes below:

> I choose Group C because for students in Group A and in Group B the measurements for distance are not the same, which means we don't know who is right and who is wrong. Group C students will provide another view, and maybe C measurements will be similar to another group whether it is A or B. (3/24c)

> I choose B because the results are not the same, in the diagram. So I calculated my own way which disproves the diagram answers. (3/08a)

> The reason why I have chosen the calculation of Group A: if you calculate all answers and divide it by 5 you get 433.02. The difference here is only 2, but they left that out, the 2. So I agree with A. (5/05a)

> I choose B because I agree [with] its average. We've calculate again. (4/15a)

> My choice is B, because all the readings in B are more closely matched than in Group A. They vary too much in Group A. (2/07a)

Level 2 argumentation. A large number of students provided a claim simply based on the data provided. The following responses are typical:

> I agree with A. If you look at the average of the distance in Group A, it does not differ that much from Group B, so that's why I say their results agree. (5/03c)

> I choose B: No, our results don't agree, because the results of column A are low than B, because A is 433.0 and B 434.8. That's why I choose B. The results of A and B are not equal. (3/07a)

> No the results are not the same because the average is not the same and the distance (mm) are not exactly the same. From release number 1 to release number 4 are not agreed each of them. (4/03c)

A number of other students based their claim on a warrant, often expressed as a theoretical statement that the results *should* or *should not* agree. This is often accompanied by an explanatory backing:

> I think the results agree, because you will seldom get the same results each time you repeat an experiment. The ball could roll in a (slightly) different direction at a different speed and end up landing in various places. There could be quite a variety in distances although they are more-or-less in the same region of distance. (2/01a)

> I chose C because the differences between the averages are because of the new theory that mathematicians have found, called the "chaos theory", which means that although the height h is the same, the ball released could be at a different speed. It could hit an irregularity in the wood, which would affect where it lands. (2/05d)

Several students use an argument with a claim, no explicit warrant but with backings similarly focusing on the procedures of experimentation:

> I think the results agree. There are different factors that can affect the distance of the ball when it lands, such as wind etc. (2/12a)

> The results do not agree, because the landing of the ball depends on the way you put the ball on the height. If you put it, in the same time pushing, it will land far, but if you just let it will land close. (3/20b)

> I agree with A. When the first group releases the ball it will land on distance d and because of the special paper the ball will leave a dent or mark. So when the other groups release the ball it might land near the spot where the previous group's ball landed (2 or 3 mm away) but because of the mark the previous ball left, the ball will slightly roll into the dent of the previous group's ball. (1/02d)

Level 3 and Level 4 argumentation. Arguments at Level 3 include a hint of a rebuttal. Often this mode emerged from students selecting option C (I have another view) as illustrated below:

> I think it is C. If we look at the averages they seem fairly equal but what makes up those averages? Distance *d* in mm at each release makes up the average but are we accurate at each release? Each time the ball hits the ground there must be change in the pressure that is inside (air) which means that the test is not fair, as probably with the change of the ball's pressure there will be a change in weight influencing how far it will go. This theory might be true if we are dealing with an air inflated ball. (1/01a)

The student opts for C in order to rebut a warrant for claim A, that is the results are the same as the averages are close together. This warrant is being rebutted by drawing attention to the changes in pressure in the ball for the repeated readings thus amounting to a set of "unfair" trials. The "pressure" warrant is qualified towards the end of the quote.

Arguments at Level 4 contain a clearly identified rebuttal. The quote below includes a claim (B), followed by a warrant and qualifier, then a rebuttal, possibly even two, with reference to the data, closing with a second warrant:

> I agree with B because the distance will always be different, although it won't differ by a great amount. The average is almost the same but it won't be exactly the same. You may get a drop distance similar, such as Group A's drop 1 and Group B's drop 2. And the measurement would be slightly out which would also affect it. (2/09a)

Some Factors Affecting Argumentation

The relationship between the initial argumentation levels and school resources is indicated in Table 2. As suggested earlier, school resource levels in South Africa may be a short-hand for a cluster of contextual factors, such as teacher qualifications and experience, school management effectiveness, facilities for engaging in practical work, the social-economic status of the learners, their linguistic ability in the language of teaching and their cultural background. Some of these contextual factors are directly recognisable in the data. For instance, responses provide an insight into how the level of argumentation on interpreting experimental data is influenced by the exposure to practical work. For instance, the reference of balls all rolling into a hollow (as in the last Level 2 quotation above, that is student 1/02d) may well be based on experience with a standard school investigation exploring the impressions balls make in sand when dropped from different heights. In addition, a few students said they resolved the argument by *doing* the "Rolling Ball" task as part of the subsequent group discussion and based their argument on their own findings.

The different quotations show that the mastery of terminology specifically related to the topic of measurement is also a crucial factor. Such mastery is often determined by school experience. For instance, only in one school, students used (or overused) the term "fair testing". In several other schools, the verb "to calculate" was used by many students for any manipulation to do with numbers: adding, dividing and even recording experimental readings. The suggestion "to re-calculate" then meant "to collect a new set of measurements". Across schools, terms like *accurate* or *precise* measurements are rarely used, and if so, they have a very general meaning as in *good* or *acceptable* measurements.

A further factor affecting levels of argumentation seems to be the familiarity with the function of diagrams. When presented with the sketch and the readings in the task some students disputed the readings as they attempted to measure the actual distance *d* on the task sheet. Several others based their argument on comparing the reading they found of the distance in the diagram with the numerical readings provided.

Changes in the Use of Argumentation Levels after Discussion

Students recorded their individual arguments, discussed these within their groups, and recorded their final claim with its justification after this discussion. They were also asked to explain why they did not support the alternative claims. A large number of groups (41 of 70 groups) completed only one post-discussion response sheet between them. This approach is clearly related to school routines as at two schools all or almost all groups reported this way. These schools were distributed across the resource spectrum. For subsequent analysis purposes, such a common response was attributed to each group member. This strategy may affect the validity of the interpretation of the changes in the individual levels of argumentation. Table 3 summarises for individual students the level of argumentation before and after the group discussion. The 20 students (i.e. five groups) who did not write a post-discussion response have been included in the category of unclassifiables.

The data in the last column and in the bottom row of the table reflect the totals for each level of argumentation before and after the group discussions, respectively. Thus, Table 3 shows that for the whole sample the use of Levels 3 and 4 has

Table 3. Changes in the use of levels of argumentation after group discussions

		Pre-discussion levels of argumentation					
		Level 1	Level 2	Level 3	Level 4	Unclassifiable	Total
Post-discussion levels of argumentation	Level 1	46	6	—	—	7	59 (22%)
	Level 2	16	88	1	1	8	114 (43%)
	Level 3	3	14	1	—	2	20 (8%)
	Level 4	4	7	1	2	—	14 (5%)
	Unclassifiables	8	30	—	1	20	59 (22%)
	Total	77 (29%)	145 (55%)	3 (1%)	4 (1%)	37 (14%)	266 (100%)

increased after the group discussion from 2% (1 + 1) to 13% (5 + 8) of all students in the sample. The proportion of unclassifiable responses has risen to a disconcerting proportion (22%) following discussion, mainly due to the 20 non-responses.

The remaining columns and rows in Table 3 represent changes in argumentation levels for individual students. If we disregard the unclassifiable responses, a total of 190 students provided codeable pre- and post-discussion responses. The shaded diagonal indicates those students who used the same level of argumentation before and after the discussion. The area above the diagonal includes the students who decreased their level of argumentation, that is 8 (6+1+1) students. The area below the diagonal includes those that increased their level of argumentation after the group discussion, that is a total of 45 (16+3+14+4+7+1) students. In summary 72% of these students (137/190) continued using the same level of argumentation, 24% of students improved (45/190) and 4% regressed (8/190). A Mann-Whitney statistical analysis of these data confirms a highly significant difference between the pre- and post-discussion results with $p < 0.01$ ($U = 21181$, $n_1 = n_2 = 190$).

Factors Affecting the Change in the Ability to Argue

Generally, changes in argumentation after the group discussion fall into three types. The first type shows no change in the level of the argument. Students show no real engagement with the evidence or with peers' arguments, but instead attempt to seek an explanation for *why* the differences of the mean have occurred. The second type of change includes a selection of one of the arguments provided by group members. The last type of change after the group discussion shows a new argument created from components of the original individual arguments.

The levels of argumentation after the small group discussion are considerably influenced by the students' perception of the purpose of the group discussion. These perceptions seem to differ from school to school. About two out of three groups, particularly those from less-resourced schools, considered that the discussion should lead to a commonly agreed argument, mostly recorded by a designated scribe. For instance, 10 (out of 70) groups saw the purpose as reaching an agreement on choice A, B or C, so they reported only post-discussion claims, but no warrants, data, backings and rebuttals. Several of the group members formulated this as "I was right!" or "X was right as usual". In all but two cases, this happened in all-male groups.

Several selection strategies can be inferred when the initial arguments of individual group members differ. Firstly, written comments show that five groups used a democratic approach, that is "the majority counts" (4/12c), to select the group argument from the various individual arguments. Clearly, this approach will only pay minimal attention to minority views, let alone to seeking evidence for rebuttals or counter claims. Six groups use a second strategy to construct a common view: they identify the argument of an "authoritative" group member, who simply explains his/her position, that is "I have chosen X's answer because he is always right" (3/01a). This strategy was more common amongst all-male than all-female groups. Five groups (out of 70) could not reach a common stand and did not provide any post-discussion arguments. All of these are all-female groups. Lastly, about a third of the groups make a judgement about the quality of the individual argument and the one perceived as most sophisticated is then adopted. Mostly, the quality judgements opt for the highest level of argument and therefore group members improve their level of argumentation. Less than a quarter of the groups go beyond selecting an argument proposed by an individual group member, but create an amalgamation of several of these arguments.

Two methods are used to build consensus and avoid confrontation. Apart from "the majority counts" strategy mentioned above, a "cop-out" method is used for avoiding taking any position about the choices and argument of group members. After the discussion, several groups opt for the view of Group C ("I have another idea which I will explain"). They suggest that Group C should take another set of measurements and that these (still unknown) measurements should settle the argument if the sets of readings were similar or not—a "referee view".

Discussion and Conclusion

Methodological Issues

The study raises a number of methodological issues around the argumentation task we used. Firstly, one could argue that argumentation levels will depend on the conceptual understanding of the evidence to be used (Maloney & Simon, 2006). In our task students were required to construct an argument about the interpretation of experimental data provided by the task, calling on their understandings of measurement, that is their understanding of concepts of evidence (Gott & Duggan, 1996). Such understanding of the concepts of evidence may change with experience of practical work. For instance, Kanari and Millar (2004) have shown that the appreciation of non-covariation of variables improves with hands-on practical experiences, and Lubben et al. (2001) report similar improvement for the understanding of repeatability in measurement. Thus, differences in students' understanding of concepts of evidence may explain some of the difference in argumentation levels for students in less- versus better-resourced schools. More specifically, Jiménez-Aleixandre et al. (2000) have shown that the pattern of argumentation based on unquestioned data provided by the task differs from that based on data collected by the students themselves. Other types of science-based argumentation tasks draw on evidence from declarative knowledge, for example the task "Is Euglena a plant or an animal, neither or both?", as described by Osborne et al. (2004) uses knowledge of classification of organisms. One would expect that performance in such declarative argumentation tasks may be less dependent on the resource levels of schools, and thus that these argumentation tasks will provide fairer and more valid assessment instruments.

Secondly, the structure of our task is semi-closed as it provides the possible—be it very general—claims in the options presented by the cartoon figures. This format has been used frequently (Keogh & Naylor, 1999, 2007). Although our options include an alternative such as "I have another view which I will explain", a response with a claim only (Level 1) requires a selection of one of the options provided, and thus hardly deserves credit as an argument. An alternative way for students to display a range of argumentation abilities may be provided by a two-phase argumentation task—an open-ended first phase, and a choice of claims as a second phase. An example of such a structure is provided by Jiménez-Aleixandre et al. (2000) who present a scenario describing the colour differences in the feathers of farmed and wild chickens. In the first, open-ended part, participants had to argue why the colour of the feathers of farm chickens was different to those of the same species in the wild. The second task component was closed, where participants were presented with a number of different possible explanations (claims) as to what might have caused the different colours of feathers and they were required to discuss the merits/demerits of each claim. However, the effective use of an open-ended phase is dependent on students' familiarity with the terminology of the argumentation topic. If unfamiliar with the issues, in this case of measurement, a task providing competing claims may be the only way.

A third methodological issue relates to the unit of analysis for this study, that is the individual student. The validity of the inferences from the written data representing the initial argumentation level may be affected by the students' ability to communicate their arguments in writing. After the group discussion, more than half of the groups returned only one response sheet. The attribution of this common response to each group member may overestimate the changes in the individual levels of argumentation. However, if the subset of individual pre- and post-responses are considered only (n = 85), the pre–post changes in argumentation levels show the same proportions and trends as for the total sample. This aspect of the analysis, however, may reduce the validity and affect the robustness of the findings of this study.

Findings about Students' Ability of Argumentation

Acknowledging the caveats above, the study provides some important findings. The data show that before the group discussion more than four out of five Grade 10 students use an Argumentation Level 1 or 2. The vast majority of the other students provide unclassifiable argumentation responses. Progression onto higher levels depends on the inclusion of rebuttals, which in every day argumentation is not so prominent: the everyday nature of argument focuses on assembling *supporting* evidence for a claim, rather than *dislodging* claims or backings (Andrews, 1995).

The findings suggest a relationship between the quality of students' argumentation and the resource level of their school. On a very practical point, the data show that a very small number of students probably have access to equipment, have experience with taking sets of measurements and are at liberty to set up an experiment to "test out an idea". As shown by Kanari and Millar (2004) the generation of experimental data will refine students' understanding of the validity and reliability of the measurements, and this in turn could stimulate rebuttals, backings and qualifiers, thus increasing the level of the argumentation. In contrast, lack of experience with practical work may well focus students on correct answers (Tsai, 1999), rather than draw on skills for using experimental evidence, the central demand of this task.

In turn, within the South African context the resource level of a school may be an indicator of a more complex set of contextual characteristics influencing argumentation levels. In this respect, Khan (2004) makes the point that language achievement is a reliable *proxy* for a school's resource levels. If this is accepted, one may assume that poorer resources may not, in themselves, provide a handicap for the development of the ability to argue, but may simply be indicative of language difficulties. Indeed, several quotations illustrate the point that some students struggle to express sometimes quite sophisticated ideas in written English. One can therefore expect that versatility in the language of instruction may well increase the validity of the communicated level of argumentation, as it is likely to reflect more closely the cognitive level of argumentation of the individual student. However, language proficiency is not a necessary pre-requisite of high argumentation levels *per se*. If small group discussions are used as learning opportunities in a social context (Lemke, 1997; Mortimer & Scott, 2003), language proficiency may become a facilitator for improvement of the level of argumentation. However, such learning based on group interaction does not need to consist of exchanges in the language of instruction. Code switching for second language learners (Rollnick, 2000), and even conversations in the mother tongue would need to be encouraged, by structured tasks providing relevant terminology in the mother tongue and even provide bilingual evidence on which the argument is meant to draw.

The lower levels of argumentation in poorer resourced schools may also be attributed to strong cultural influences, in that students in these schools often come from African cultural backgrounds. A study by Scholtz et al. (2008) of argumentation patterns of teachers with an African cultural background suggests that they avoid rebuttals and use alternative moves in their arguments, such as questions, selective summaries and critical reference to the task. The authors note the frequent use of the word "we" and the provision of new claims, warrants and backing as parallel contributions rather than opposing those of others. Even if opposing evidence is suggested, it is often preceded by a summary of agreed views. Posing questions, referring to a (neutral) authority such as the task description and introducing qualifiers, are used as non-confrontational ways of disagreement. The participants listen to and engage with the comments made by their discussion partners, which differs from competitive principles often prevalent in argumentation discussions (Forman, Larreamendy-Joerns, Stein, & Brown, 1998). The extent of the verbal interaction amongst the students could not be ascertained as we relied on their written responses. However, the fact that more than half of the groups returned only one response sheet might indicate a desire for consensus rather than conflict.

Similar cultural influences on argumentation conventions are reported by Hudicourt-Barnes (2003) and Liu (2005), who have shown that the formats of argumentation used by Caribbean and Chinese students respectively put less emphasis on confrontation of claims or supporting backings put forward by peers. Such a cultural avoidance of rebuttals allows little scope for the higher levels of argumentation in the scheme developed by Erduran et al. (2004). One implication is that a culturally sensitive assessment scheme should de-emphasise explicit rebuttals and highlight two additional aspects. First, more differentiation at the upper assessment levels should be based on the quality and explicitness of warrants and backings. Andrews (2007) suggests that facilitators of learning should "recognise the range of argumentative forms and practices" (p. 6) across cultures, exemplifying that "in the Confucian-heritage cultures of East-Asia exposition and clarity is a highly prized quality of thinking" and that students from these cultures often "find it difficult to adjust to European cultural demands for a critical dimension" (p. 11) in argumentation. A requirement for clarity of evidence equally forces students into elaboration, explanation and justification of their assertions, that is the purpose of argumentation. We do not suggest a return to the linear argument (Harris, 1998) and a disregard of contradictory and alternative evidence. However, different expressions of disagreement need to be credited within a culturally sensitive assessment scheme. For instance, Table 4 shows a modified version of the "culturally inclusive" assessment scheme proposed by Scholtz et al. (2008). The inclusive descriptors are shown in italics.

In assessing the levels of argumentation, the prominence of the use of a rebuttal for arguments beyond Level 2 has been reduced, and instead the scheme highlights the use of qualifiers, of questions querying the relevance or quality of the evidence, warrants or backings for a stated claim, and of the method of pros-and-cons with a reasoned on-balance summary claim. All these aspects feature prominently in scientific reasoning. They link assertions with the evidence but are less confrontational than conventional rebuttals, and thus culturally more acceptable. Qualifiers delineate the applicability of the claim and thus make it more precise. Probing questions suggest a critical interrogation of the claims, warrants and backings offered by

Table 4. Levels of inclusive argumentation (modified from Scholtz et al., 2008)

Levels	Description
1	Arguments that are a simple claim versus a counter-claim or a claim versus a claim
2	Arguments consisting of a claim versus a claim with either data, warrants or backings but no rebuttals
3	Arguments with a series of claims or counter-claims with either data, warrants or backings with the occasional weak rebuttal *Arguments with a series of claims with data and an explicit reasoned qualifier*
4	Arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counterclaims *Arguments with a claim with a reasoned question querying the warrants, backings or use of the data*
5	Displays an extended argument with claims supported by data and warrants, with more than one rebuttal *Displays an extended argument with claims and counter claims each supported by data, warrants/backings, with a reasoned summary claim*

others. An elaboration of claims and counter claims (each with supporting data and warrants/backings) together with a summary claim emerging from several claims and counter claims, constructs a web of assertions and weighted evidence.

The findings suggest there is a gender effect in the ways small group discussions deal with conflicting individual arguments. All-male groups are more likely to adopt an argument from an "authority" or see the discussion as a competition of views (and brag about having "won"). In contrast, all-female groups are more likely to use a "democratic" approach or avoid forcing a common stand by not reporting any post-discussion argument. These gender-related strategies mirror those in the study by Tolmie and Howe (1993) where all-female groups seek for the common aspects of individual views, and all-male groups generally are upfront about differences in views which are seen as competing. In a review of research on gender aspects of group discussion strategies Scanlon (2000) confirms that all-female groups are mainly concerned with maintaining good relationships between group members. This suggests that all female groups will put a considerable emphasis on the identification of qualifiers and the construction of parallel arguments. As long as these efforts are based on interacting with the evidence, they should also gain merit in an adapted assessment scheme.

The data show that small group discussions, without any further scaffolding supporting the argumentation, improve the argumentation level of about a quarter of the students in the sample. The implication seems to be that even without extensive teacher professional development efforts as illustrated by Simon, Erduran, and Osborne (2006), carefully structured small group discussions *per se* provide opportunities for students to improve their argumentation skills. The data also indicate that the positive impact of small group discussions may be increased by clarifying the role of group discussions in argumentation as sense making of evidence, rather than an attempt to come to a common stand on the claim, or even on the warrants and backings. Studies of African classrooms where teachers attempted to make the structure and expectations of argumentation explicit to students before they engaged in scientific argumentation show that this can be a productive strategy (Braund, 2007). Webb, Williams, and Meiring (2008) investigated the use of concept cartoons and writing frames over time, to scaffold learners' thinking. They suggest that the approach is fully appropriate within the context of the South African curriculum. Their findings seem to hint at progression in learner argument over time. Thus, some meta-cognitive aspects will need to be included in the introduction of argumentation lessons in order to provide an experience of different purposes of small group discussions, including the specific purpose of argumentation discussions.

References

Abd-El-Khalick, F., Lederman, N. G., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire (VNOS): Towards valid and meaningful assessment of learners' conceptions of the nature of science. *Journal of Research in Science Teaching*, *39*(6), 497–521.

Andrews, R. (1995). *Teaching and learning argument.* London: Cassell.

Andrews, R. (2007). Argumentation, critical thinking and the postgraduate dissertation. *Educational Review, 59*(1), 1–18.

Braund, M., Lubben, F., Scholtz, Z., Sadeck, M., & Hodges, M. (2007). Comparing the effect of scientific and socio-scientific argumentation tasks: Lessons from South African classrooms. *School Science Review, 88*(324), 67–76.

Chan, C. (2001). Peer collaboration and discourse patterns in learning from incompatible information. *Instructional Science, 29,* 443–479.

Dagher, Z., Brickhouse, N., Shipman, H., & Letts, W. (2004). How some college students represent their understandings of the nature of scientific theories. *International Journal of Science Education, 26,* 735–755.

Department of Education (DoE). (2004). *Further education and training: G10–12 (General): Overview.* Pretoria: Department of Education.

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.

Driver, R., Asoko, H., Leach J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23,* 5–12.

Duschl, R. (2008). Quality argumentation and epistemic criteria. In S. Erduran & M. Jiménez-Aleixandre (Eds.), A*rgumentation in science education* (pp. 159–175). Dordrecht: Springer.

Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38,* 39–72.

Erduran, S., Osborne, J., & Simon, S. (2004). TAPping into argumentation: Developments in the application of Toulmin's argumentation pattern for studying science discourse. *Science Education, 88*(6), 915–933.

Facione, P., & Facione, N. (1992). *The California Critical Thinking Disposition Inventory (CCTDI).* Millbrae, CA: California Academic Press.

Finkel, E. (1996). Making sense of genetics: Students' learning in classrooms: Reframing classroom research. *Journal of Research in Science Teaching, 33,* 345–368.

Forman, A., Larreamendy-Joerns, J., Stein, M., & Brown, C. (1998). "You are going to want to find out which and prove it": Collective argumentation in a mathematics classroom. *Learning and Instruction, 8*(6), 527–548.

Glassner, A., & Schwarz, B. (2007). What stands and develops between creative and critical thinking? Argumentation? *Thinking Skills and Creativity, 2,* 10–18.

Gott, R., & Duggan, S. (1996). Practical work: Its role in the understanding of evidence in science. *International Journal of Science Education, 18,* 791–806.

Harris, R. (1998). *Introduction to creative thinking.* Mahwah, NJ: Lawrence Erlbaum.

Hogan, K. (1999). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching, 36,* 1085–1109.

Hudicourt-Barnes, J. (2003). The use of argumentation in Haitian Creole science classes. *Harvard Educational Review, 73*(1), 73–93.

Jiménez-Aleixandre, M., & Erduran, S. (2008). Argumentation in science education: An overview. In S. Erduran & M. Jiménez-Aleixandre (Eds.), A*rgumentation in science education* (pp. 3–27). Dordrecht: Springer.

Jiménez-Aleixandre, M., Rodríguez, A., & Duschl, R. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84*(6), 757–792.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748–769.

Kelly, G., & Chen, C. (1999). The sound of music: Constructing science as a socio-cultural practice through oral and written discourse. *Journal of Research in Science Teaching, 36*(8), 883–915.

Kelly, G., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education, 86*(3), 314–342.

Keogh, B., & Naylor, S. (1999). Concept cartoons, teaching and learning in science: An evaluation. *International Journal of Science Education, 21*(4), 431–446.

Keogh, B., & Naylor, S. (2007). Talking and thinking in science. *School Science Review, 88*(324), 85–90.

Khan, M. (2004). For whom the school bell tolls: Disparities in performance in senior certificate mathematics and physical science. *Perspectives in Education, 22*(10), 149–156.

Koosimile, A. (2004). Out-of-school experiences in science classes: Problems, issues and challenges in Botswana. *International Journal of Science Education, 26*(4), 483–496.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning of scientific thinking. *Science Education, 77,* 319–337.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28*(2), 16–26.

Lajoie, S., Lavigne, N., Guerrera, C., & Munsie, S. (2001). Constructing knowledge in the context of BioWorld. *Instructional Science, 29,* 155–186.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education, 25*(11), 1387–1408.

Leach, J., Driver, R., Millar, R., & Scott, P. (1997). A study of the progression in learning about "the nature of science": Issues of conceptualization and methodology. *International Journal of Science Education, 19,* 147–166.

Lemke, J. (1997). Cognition, context and learning: A social semiotic perspective. In D. Kirschner & P. Whitson (Eds.), *Situated cognition: Social, semiotic and psychological perspectives* (pp. 37–46). Mahweh, NJ: Erlbaum.

Liu, L. (2005). Rhetorical education through writing instruction across cultures: A comparative analysis of select on-line instructional materials on argumentative writing. *Journal of Second Language Writing, 14*(1), 1–18.

Lubben, F., Campbell, B., Buffler, A., & Allie, S. (2001). Point and set reasoners in practical science measurement by entering university freshmen. *Science Education, 85*(4), 311–327.

Maloney, J., & Simon, S. (2006). Mapping children's discussions of evidence in science to assess collaboration and argumentation. *International Journal of Science Education, 28*(15), 1817–1841.

Mercer, N., Dawes, L., Wegerif, W., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal, 30*(3), 359–377.

Mortimer, E., & Scott, P. (2003). *Meaning making in secondary science classrooms.* Buckingham: Open University Press.

Nsamenang, B. (1999). *An image of the self in African social thought.* Newberry Park, CA: Sage.

Osana, H., & Seymour, J. (2004). Critical thinking in pre-service teachers: A rubric for evaluating argumentation and statistical reasoning. *Educational Research and Evaluation, 10*(4), 473–498.

Osborne, J., Erduran, S., & Simon, S. (2004). *Ideas, evidence and argument in science (IDEAS PROJECT).* London: Kings College, University of London.

Paul, R. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world.* Santa Rosa, CA: Foundation for Critical Thinking.

Paul, R., & Elder, L. (2004). *The miniature guide to critical thinking concepts and tools.* Dillon Beach, CA: Foundation for Critical Thinking.

Richmond, G., & Striley, J. (1996). Making meaning in classrooms: Social processes in small group discourse and scientific knowledge building. *Journal of Research in Science Teaching, 33,* 839–858.

Rollnick, M. (2000). Current issues and perspectives on second language learning in science. *Studies in Science Education, 35,* 93–122.

Ryder, J., Leach, J., & Driver, R. (1999). Undergraduate science students' images of science. *Journal of Research in Science Teaching, 36,* 201–219.

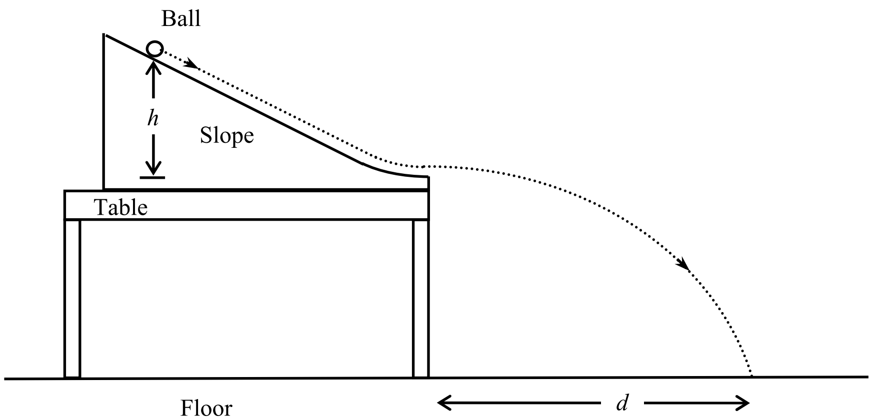
Sampson, V., & Clark, D. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*(3), 447–472.

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education, 93*(3), 448–484.
Sandoval, W., & Millwood, K. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23–55.
Sandoval, W., & Reiser, B. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry, *Science Education, 88*(3), 345–372.
Scanlon, E. (2000). How gender influences learners working collaboratively with science simulations. *Learning and Instruction, 10,* 463–481.
Scholtz, Z., Braund, M., Hodges, M., Koopman, R., & Lubben, F. (2008). South African teachers' ability to argue: The emergence of inclusive argumentation. *International Journal of Educational Development, 28*(1), 21–34.
Schwarz, B., Neuman, Y., Gil, J., & Ilya, M. (2003). Construction of collective and individual knowledge in argumentation activity. *Journal of the Learning Sciences, 12*(2), 219–256.
Sherman, G., & Klein, J. (1995). The effect 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), 235–260.
Soudien, C. (2007). The "A" factor: Coming to terms with the question of legacy in South African education. *International Journal of Educational Development, 27,* 182–193.
Steyn, J. (2000). Quality education and equality in education: A dilemma for democratic South Africa. *South African Journal of Education, 20*(1), 46–49.
Tabulawa, R. (1997). Pedagogical classroom practice and the social context: the case of Botswana. *International Journal of Educational Development, 17*(2), 189–204.
Tao, P. (2003). Eliciting and developing junior secondary students' understanding of the nature of science through a peer collaboration instruction in science stories. *International Journal of Science Education, 25,* 147–171.
Tolmie, A., & Howe, C. (1993). Gender and dialogue in secondary school physics. *Gender and Education, 5,* 191–209.
Toulmin, S. (2003). *The uses of argument* (Updated ed.) Cambridge: Cambridge University Press.
Tsai, C.-C. (1999). "Laboratory exercises help me memorize the scientific truths": A study of eight graders' scientific epistemological views and learning in laboratory activities. *Science Education, 25,* 147–171.
Tsai, C.-C., & Liu, S.-Y. (2005). Developing a multi-dimensional instrument for assessing students' epistemological views towards science. *International Journal of Science Education, 27*(13), 127–149.
Venter, E. (2004). The notion of Ubuntu and communalism in African educational discourse. *Studies in Philosophy of Education, 23,* 149–160.
Webb, P., Williams, Y., & Meiring, L. (2008). Concept cartoons and writing frames: Developing argumentation in South African science classrooms? *African Journal of Research in Mathematics, Science and Technology Education, 12*(1), 5–18.
Wray, D., & Lewis, M. (1997). *Extending literacy: Children reading and writing non-fiction.* London: Routledge.
Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*(1), 35–62.

Appendix. Laboratory Situation: Rolling balls

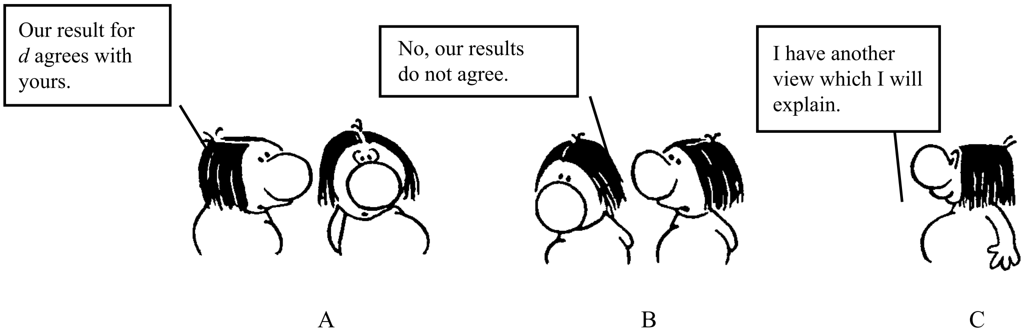
Some students do an experiment in the laboratory. They clamp a wooden slope near the edge of a table. They release a ball from a height *h* above the table as shown in the diagram. The ball leaves the slope *horizontally* and lands on the floor a distance *d* from the edge of the table. Special paper is placed on the floor on which the ball makes a small mark when it lands.

The students want to investigate how the distance *d* on the floor changes when they vary the height *h*. They use a metre stick to measure *d* and *h*.



Two groups of students compare their results for *d*. Their readings for five releases from the same height (90 mm) are shown below, together with their averages.

	Group A	Group B
Release	*d* (mm)	*d* (mm)
1	439.5	435.4
2	438.3	439.2
3	433.1	428.0
4	422.8	433.1
5	431.3	438.3
Average:	**433.0**	**434.8**



Task sheet: Rolling balls

Your name is: ____

You have been asked to work with three other students.
What are the names of the other people in your group?
1. _________________ 2. _____________________ 3. _________________

Read about the laboratory situation on the previous page.
Before any group discussion what is the option you agree with? A/B/C (circle one)
Explain your choice:

____
____
____
____
____

Now discuss your options and reasons in the group. Listen to the arguments of the others and try to respond.
After the discussion and hearing the arguments of the students in your group, write down your final choice: A/B/C (circle one only)

Explain the reasons for your choice and why you would not go for the other options.

____
____
____
____
____
____