E–Learning and Digital Media
Volume 10, Number 4, 2013
www.wwwords.co.uk/ELEA

Science in Writing: learning scientific argument in principle and practice

BILL COPE, MARY KALANTZIS & FOUAD ABD-EL-KHALICK
University of Illinois at Urbana-Champaign, USA
ELIZABETH BAGLEY
University of Wisconsin-Madison, USA

ABSTRACT This article explores the processes of writing in science and in particular the 'complex performance' of writing a scientific argument. The article explores in general terms the nature of scientific argumentation in which the author-scientist makes claims, provides evidence to support these claims, and develops chains of scientific reasoning to coordinate claims and evidence. The article then describes a case in which two classes of Grade 8 students in a New York City school wrote scientific arguments in a web-writing and peer review environment that provides each writer with 'as-you-go' formative assessment on the constitutive elements of their arguments.

Introduction

The Common Core Standards for Writing encompass a writing genre called 'argument' intended for use across the curriculum, including the writing of science. This article explores the notion of scientific argument as a form of academic writing, and its assessment, in theory as well as in classroom practice.

The grounding for our analysis is data emerging from a cluster of research and development projects at the University of Illinois funded by the Institute of Education Sciences, US Department of Education. As part of this research and development work, we have designed, developed and trialed a web-writing and formative assessment environment, *Scholar*. The student argument writing and peer review processes described later in this article occurred within *Scholar*.[1]

A Perceived Crisis in Science Learning

Considerable concern has been expressed in recent years about school science education in the United States. In public discourse, this has translated into a perception that there is a crisis in science learning. Supporting this perception, the Science and Engineering Indicators 2010 Report (National Science Board, 2010) highlights the need for improved science education in the USA, amidst international comparisons of student achievement in science placing the USA at the average score. Although the 2007 Trends in International Math and Science Study (TIMSS) shows a slight improvement in US student science scores since 1995, the USA still ranks 10 out of 17 of the participating developed nations (Gonzales et al, 2008). Further, although the 2009 Program for International Student Assessment (PISA) average science literacy score of 15-year-olds in the USA was not measurably different from the Organization for Economic Co-operation and Development (OECD) average score, the US score was lower than the average score in 12 of the 33 other OECD countries (OECD, 2010).

For a country that has historically prided itself in being a world leader in science and technology and that has managed, albeit through significant dependence on international talent, to

 http://dx.doi.org/10.2304/elea.2013.10.4.420

continue its leadership in scientific and engineering knowledge production, these are disappointing outcomes that represent a setback over a period of several decades. This is particularly the case when it is considered that these international assessments purport to focus on deep scientific understanding, rather than memorization of facts and theories (National Research Council, 2011).

This is the context for two major new standards initiatives in the United States, the Next Generation Science Standards (NGSS) and Common Core State Standards (CCSS) for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects (Common Core State Standards Initiative, 2010; Next Generation Science Standards, 2012). These standards represent a fundamental shift in education policy and practice. Both initiatives replace highly variable standards from state to state with consistent standards that are adopted by most states. They also represent a paradigm shift away from the memorization of content knowledge towards a focus on developing disciplinary orientations and broad epistemic capacities. Both CCSS and NGSS focus on an underlying capacity to do the kinds of knowledge work characteristic of academic disciplines and applied disciplinary knowledge. Nowhere is this shift clearer than in the newly emphasized area of writing. The CCSS stress the role of writing in almost every subject area, where it is used as a vehicle to represent disciplinary knowledge and demonstrate higher-order thinking skills. These standards focus on three canonical types of written text: argument, information and explanation, and narrative. The first two of these text types are particularly important to science.

For instance, at the Grade 8 level, the CCSS specify that students will be able to write arguments to support claims with clear reasons and relevant evidence. The relevant performance outcomes state that students must be able to:

> a. Introduce claim(s), acknowledge and distinguish the claim(s) from alternate or opposing claims, and organize the reasons and evidence logically.
>
> b. Support claim(s) with logical reasoning and relevant evidence, using accurate, credible sources and demonstrating an understanding of the topic or text.
>
> c. Use words, phrases, and clauses to create cohesion and clarify the relationships among claim(s), counter-claims, reasons, and evidence.
>
> d. Establish and maintain a formal style.
>
> e. Provide a concluding statement or section that follows from and supports the argument presented. (Common Core State Standards Initiative, 2010, p. 42)

Doing science in this way is very different from our historical practices of content learning that focus on memorizing facts, definitions of concepts, and the application of formulae from which it is possible to deduce unequivocal answers. Science in the new standards becomes a process of reasoning in which claims can be questioned, evidence interrogated, and reasoning analyzed. Formerly a body of straightforward empirical information, science becomes a form of reasoning from evidence, a way of doing and thinking.

The Practice of Science

This enormously significant shift in pedagogical focus is designed in part to reflect more faithfully the social and cognitive practices of science. Doing 'real' science involves higher-order thinking (Etkina et al, 2005; Mestre et al, 2009) and evidential reasoning (Cetina, 1999). It involves a certain kind of discourse, or 'talking science' (Halliday & Martin, 1993; Halliday, 2004) by 'observing, describing, theorizing, questioning, challenging, arguing, designing experiments, following procedures, judging, evaluating, deciding, concluding, generalizing, reporting' (Lemke, 1990, p. ix). It includes constant reflection on 'the epistemological bases of the discipline' (Grotzer, 2009, p. 61). It involves the activities and skills needed to develop a coherent, empirically based scientific argument (Driver et al, 2000; Osborne, 2005; Abi-El-Mona & Abd-El-Khalick, 2006; Jimenez-Aleixandre & Eurduran, 2008; Newell et al, 2011). Doing science involves thinking and acting in the characteristic ways of a scientist in a context of situated cognition (Latour & Woolgar, 1986; Gee, 2004a; Gee, 2004b). It also entails working in a community of practice (Lave & Wenger, 1991; Wenger et al, 2002). The purpose of these standards is not (mainly) to nurture a new generation of scientists, but to create an informed citizenry capable of everyday 'science literacy'.

How does one translate into classroom practice the pedagogical means that will enable teachers to structure learners' experiences in ways that are more true to scientific practice and the broader social purposes of scientific literacy? In this article, we attempt to answer this question in the particular case of writing scientific arguments. This is a process in which students act and interact like scientists, building arguments collaboratively, receiving ongoing feedback, while being provided with extended and multiple opportunities to refine and improve their arguments.

Writing Science

In the research project upon which this article is based, we have designed a social, collaborative process of argument construction, based on the knowledge formulation and peer review processes that characterize 'real' science. A major pitfall in common pedagogical practices in the area of science argumentation is the fact that students mostly work on their own to construct arguments with little opportunity for reflection (e.g. during one class period). These arguments are then evaluated by an external authority, namely the teacher. This approach is strikingly different from the way in which scientists work and build arguments. Scientific arguments mostly are formulated within collaborative teams over a period of time and vetted through several rounds and levels of feedback and revision (conference presentations, peer review, etc.).

We have also developed a model of science argumentation (Table I) based on the Common Core Standards for Science argumentation (Common Core State Standards Initiative, 2010), the Next Generation Science Standards (Next Generation Science Standards, 2012), and synthesis of the literature on argument structure (Toulmin, 2003; Van Eemeren & Grootendorst, 2004). We have refined and reworked this model through an iterative process of design-implementation-redesign in middle-school classrooms, including the case study we describe later in this article.

Opening argument	*Thesis* – introducing the argument, and the question or problem that the argument sets out to address and the solution it sets out to find. The argument also needs to attend to audience – establishing a relationship of informed expertise (*the writer*) between the writer and the inquiring but as-yet less knowledgeable reader.
Claim 1	*Claim* – a statement that answers the original question or problem, followed by: 1. *Evidence* – appropriate and sufficient scientific data that supports the claim. 2. *Reasoning* – a justification that connects the evidence to the claim and shows why the data counts as evidence by applying scientific principles (often labeled 'warrant').
Claim 2	Next claim, as previous
Counter-claims	*Counter-claims* – an awareness of the existence of alternate or opposing claims, and a capacity to critically evaluate the evidence offered to support these claims, and the reasoning provided.
Conclusions	*Evaluation* – overall judgment, by way of conclusion.

Table I. Science argumentation model.

How then, do we characterize the discourse of science of which scientific argumentation is an instance? 'Science', say Latour and Woolgar, is 'the process whereby an ordered account is fabricated from disorder and chaos' (1986, p. 41). This may be considered an overstatement, because everyday life is not so chaotic that we cannot make sense of it. The words of our language order everyday experience quite effectively; our language helps us learn to know, at the very least, elementary scientific propositions, such as the one that dogs and cats are kinds of animals. However, in the nature of the 'natural language' that we use in everyday life, there is often imprecision and ambiguity, which scientific language cannot afford because it seeks to know the world in a more carefully focused way (Bowker & Star, 2000). Scientific language generates deeper insights than can be provided to us in the impressionistic language of everyday life. For instance, scientific argument involves chains of evidential reasoning that systematically connect documentation of concrete, empirical realities (evidence) with claims, and claims tie theses to overall judgments (Abi-El-Mona & Abd-El-Khalick, 2006). Scientific discourse also requires learners to monitor their own thinking, always self-questioning veracity and identifying possible fallacies in the meanings or sense being made. Does this evidence in fact support a claim? Does the reasoning constitute sound science?

In the sentence-by-sentence specifics of the written discourse of science (called 'formal style' in the Common Core Standards), linguists have attempted to analyze some of the peculiar aspects of the language in which scientific knowledge is packaged. By way of example, Gee (2004a) offers these two sentences, the first typical of everyday speaking; the second typical of academic writing:

1. Hornworms sure vary a lot in how well they grow.

2. Hornworm growth exhibits a significant amount of variation.

Here are some changes that have occurred in the transition from the first sentence to the second. The verbs 'vary' and 'grow' have been transformed into nouns – a process called 'nominalization' in which actions are turned into abstract things. Next, a verb with descriptive content ('vary') has been replaced by a generic verb of appearance, which connects abstract things ('growth' and 'variation'). The end point of a hornworm's development ('how well') has been replaced by a term related to the measures of science ('significant amount'). The first sentence tells you something of its creator's attitude to the subject ('sure'), but the second does not. In this transition, Gee concludes that:

> some things are lost – concrete things such as hornworms and empathy for them; [and] changes and transformations as dynamic, ongoing processes ... Some things gained are abstract ideas and relations among them; traits and the quantification and categorisation of traits; and evaluation from within a specialised domain. (Gee, 2004a, pp. 16-17)

By doing science, students learn its literacies and the ways of thinking embedded in these literacies. Lemke comes to similar conclusions about the language of science:

> There is a lot of use of the passive voice, of abstract nouns in place of verbs, of verbs of abstract relation (e.g. be, have, represent) in place of verbs of material action. It also has its preferred figures of speech, like analogy, and rhetorical patterns (e.g. Thesis-Evidence-Conclusion). It also works through a variety of activity structures, whether triadic dialogue, ordinary question-and-answer, lecture, or summary monologues, or many others. It even has its own special forms of written texts: laboratory notes, reports of experiments, theoretical treatises, and so on. (Lemke, 1990, p. 21)

Such are the representational and communicative raw materials of science and learning science. 'Science,' Lemke says, 'is the great enterprise of paying attention to the kinds of meanings that require us to go beyond natural language' (Lemke, 2004, p. 34). This is also how science becomes something bigger than one's own experiences, perspectives and voice. It becomes a body of social knowledge institutionalized and 'objectified' in the shared meanings communicated in scientific texts.

This leads us into foundational, psycholinguistic arguments about the role of formal writing in the development of scientific thinking. In Vygotsky's (1986) view, 'the development of writing does not repeat the developmental history of speaking.' Even the 'minimal development' of writing 'requires a high level of abstraction'. The motives for writing, moreover, are far from 'the 'authentic' or 'natural' ways in which children's oral language in their first years expresses their immediate needs and interests. Writing is 'more abstract, more intellectualized, further removed from immediate needs'. This detachment goes as deep as the discrepancy between thinking in the language of 'inner speech' (imported as a tool for thinking from the oral language learned as a child) and the language of writing:

> Inner speech is almost entirely predicative because the situation, the subject of thought, is always known to the thinker. Written speech, on the contrary, must explain the situation fully in order to be intelligible. The change from maximally compact inner speech to maximally detailed written speech requires what might be called deliberate semantics – deliberate structuring of the web of meaning.

Children start to write long before they have the cognitive capacity to employ the full resource of a 'deliberate semantics'. At its highest levels, this resource includes planning and drafting. However, the fact that writing instruction starts when the cognitive capacities upon which writing is ultimately based are only at a rudimentary stage of development is evidence for Vygotsky that cognitive capacity does not precede instruction, but 'unfolds in a continuous interaction with the contributions of instruction,' the 'abstract, deliberate' forms of thinking through language that are peculiar to writing (Vygotsky, 1986, pp. 98, 99, 100, 144, 101, 100).

These emerging forms of thinking are then imported back into adult speaking, and particularly 'speaking-like-writing'. This is reflected both in the intricacies of 'formal style' in scientific writing, and the explicitly articulated structures of whole texts such as scientific argument.

As students move into scientific writing, an important space opens up for what we want to call metarepresentation, or representations about representation. In the case of scientific argument, this means explicit naming of theses, claims, evidence, reasoning and judgments. In this way, metarepresentation is a tool for uncovering the principles of order in scientific texts: how they work, who they work for, why they work for them. The aim of metarepresentation is not that students should learn definitions and rules of the generic components of scientific text (somewhat like old-fashioned grammar rules). Rather, they learn how to unpack texts by developing and applying a conceptual framework that explains the design of that text (Vygotsky, 1986, pp. 92, 91, 93.) This kind of explicitness, says Cazden (2001), constitutes one of the characteristic features of academic discourse, 'the special ways of talking expected in school'. For these reasons, a degree of explicitness about academic discourse and reasoning processes needs to be built into the instruction – in the case we are studying here, the process of written scientific argumentation.

The science of scientists, and the science of school, represent the world in peculiar ways, ways that are in some respects quite different from everyday, ordinary representations. Scientists frame their knowledge in different ways from more casual, everyday knowledge. We are going to call these 'technical' or 'academic' forms of knowledge and meaning-representation. The kinds of representation that we learn in school subjects and in further education embody ways of thinking and also create kinds of people.

Vygotsky says that the long journey of learning literacy and academic discourse eventually takes learners to another cognitive plane, that of 'reflective consciousness' and a 'new awareness when you are conscious of being conscious'. Then, scientific concepts developed through the practices of schooling are transferred to everyday concepts, changing their psychological structure from the top down. 'For the young child, to think means to recall,' says Vygotsky, 'but for the adolescent, to recall means to think. Her memory is so "logicalized" that remembering is reduced to establishing and finding logical relations' (Vygotsky, 1978, p. 51). This is how scientific writing of the kind described in this article helps learners think in new ways.

Moreover, scientific literacy performs an additional function as a support medium of knowledge and a support for learning. Not only is scientific literacy a medium for thinking, but the written scientific record also affords learners an ever-present extension of memory and personal thinking, overcoming the limitations of long-term memory. Scientific texts are a social mnemonic, a social memory outside of our individual minds that form an essential supplement to adult minds. Scientific literacy, then, is not just a capacity to write science. It is also a documentary source for which we can always reach while we are writing though a process of 'looking up' a written text and sourcing it as another's data or judgment. In these ways, scientific literacy involves working with conceptual artifacts, such that our thinking and acting capacities extend well beyond our individual brains. Scientific literacy is the stuff of accumulated, collective, distributed and essentially human intelligence. This is what Gee calls 'the social mind' (Gee, 1992). Scientific literacy is the stuff of accumulated, collective, distributed and essentially human intelligence. It is the product of scientific communities. Written scientific texts are ever-present and always-needed social extensions of our personal minds.

Knowledge Representations and Complex Performance

Writing science is also a form of what we want to characterize as 'complex performance'. We want to apply this phrase to writing science in two senses. First, writing uniquely supports the representation of complex scientific thinking. The disciplinary and knowledge practices of science are an instance of 'complex performance' which can best be represented in writing. And second, writing is itself an exceedingly complex form of epistemic performance.

One centrally important practice in the complex performance of science is reasoning through argumentation. Scientific reasoning is broadly defined as 'the practice of thinking with and about scientific knowledge' (Hogan & Fisherkeller, 2005, p. 95) and involves key epistemic processes such as coordinating claims with evidence (Kuhn, 2004) and facility with 'formal' language choices. Argumentation is broadly defined as making and justifying claims by reference to evidence and reasoning for the purpose of persuasion for a variety of purposes (scientific and social). Learning to develop scientific argumentation is crucial to conducting scientific inquiry (Driver et al, 2000) and an integral mechanism for undertaking the complex performance that is science (Jonassen & Kim, 2009) and for informing and communicating scientific knowledge to a general public. Scientific argumentation involves the ability to provide 'evidence that can be confirmed with a logical argument' (American Association for the Advancement of Science, 1993, p. 11) and to weigh evidence and consider alternatives. Reasoning and argumentation are indispensable parts of contemporary science education, bringing learners close to the social, epistemological, and discursive practices of scientific endeavors, expanding their analytical and critical thinking capacities, and enhancing their understanding of the conceptual and empirical contents of science and its social purposes (Abi-El-Mona & Abd-El-Khalick, 2010; Krajcik & Sutherland, 2010).

Research, however, reveals that students get very little practice in developing their reasoning and argumentation skills and rarely build arguments even in contexts that are conducive to argumentation such as socioscientific and inquiry contexts (Duschl, 1998; Driver et al, 2000; Duschl & Ellenbogen, 2001; Duschl & Osborne, 2002). When students do attempt to construct arguments, they often lack the abilities or skills needed to structure valid arguments, such as drawing on relevant evidence to support a certain claim, building meaningful links between claims and evidence, and recognizing fallacious argumentative schemes (Zeidler, 1997). They also lack a deep understanding of language choices, like nominalization, that would underpin the strength of their case. In sum, research in science education indicates that students have difficulties utilizing the methods, skills and language choices most valued by scientists to construct scientific arguments (Mason & Santi, 1994; Jimenez-Aleixandre et al, 1998; Patronis & Spiliotopoulou, 1999; Kolsto & Ratcliffe, 2008).

Munford and Zembal-Saul argue that the metacognitive benefit to students of learning scientific argumentation is that it presents opportunities to learn not only about science content but about science processes (Munford & Zembal-Saul, 2002; Zembal-Saul, 2005). This includes an understanding of the role of language, culture, and social interaction in the process of knowledge construction (Pontecorvo, 1987; Brown et al, 1989; Eurduran & Jimenez-Aleixandre, 2008). Explicit teaching of scientific argument and its linguistic manifestations also enhances learners' engagement with argumentative discourse by rendering learners' understanding and thinking visible, thus providing a valuable tool for reflection and assessment (Abell et al, 2000; Bell & Linn, 2000; Zembal-Saul & Land, 2002; Abell, 2006; Kuhn et al, 2008). Finally, research shows that explicit highlighting of the nature of scientific argument, and the elements that go into its construction, support the development of divergent ways of thinking (Kuhn, 1991; Kuhn, 1993) and enhancing understandings of scientific ideas (Zohar & Nemet, 2002) and their representation.

Assessing Writing in Science

What does this focus on complex performance and science writing mean for assessment? Our counterpoint to the complexities of science writing is the discrete item summative assessment that is so often used as a relatively inexpensive and technically reliable means for assessing science knowledge. Empirical correctness and the accurate application of theories are disembodied from argument. However, the question is often asked about the extent to which such forms of assessment can go beyond the memorization of facts, the identification of correct definitions, drawing axiomatic conclusions, and deducing empirically correct answers in the application of theories. Such assessments have frequently been criticized for their tendency to break science into atomized fragments, each with a definitively right answer, juxtaposed beside deceptively nearly-right but actually wrong 'distractor' answers.

We do not want to offer unqualified criticism of 'selected response' assessments, because they have a place. In fact, these forms of assessment have as-yet substantially unrealized potentials in the form of discrete-item computer-adaptive and diagnostic tests (Cope et al, 2011). However, discrete-item tests need to be supplemented systematically by a broader repertoire of assessment types. As we have argued, much of the knowledge and many of the human capabilities that are most relevant to today's workplace, community, and educational settings, are instances of 'complex performance': problem solving, coming to modulated conclusions which are beyond definite yes/no dichotomies, knowledge that only makes sense when contextualized by broader knowledge and social systems, knowledge created in social collaborations, and innovation or initiative (Darling-Hammond et al, 1995; Pellegrino et al, 2001; Baker, 2004; Mislevy, 2006b; Darling-Hammond & Wood, 2008; Wagner, 2008). These are precisely the epistemic qualities that assessment of scientific writing is designed to measure.

One of the key characteristics of the complex performance of scientific argument writing is that the assessable object cannot be produced within the framework of 'test time'. Rather it is an artifact that the student necessarily generates over a series of days, even weeks. It also focuses on different aspects of scientific knowledge than is possible in selected response or short-answer assessments.

The assessment of the complex performance of writing has several distinct advantages. It provides a less mediated view than selected response or short-answer assessments of students' actual scientific reasoning processes. It is also possible to position the assessment as incremental feedback offered during the learning process – embedded formative assessment of science argumentation, as well as cumulative data for summative and progress assessments. It also brings student work and assessment of that work closer to the authentic practice of scientists as they make their professional intervention through their participation in the world of scientific discourse. For instance, it is possible to establish peer review processes for evaluating scientific argumentation (and counter-argumentation) analogous to those that undergird scientific practice (Kalantzis & Cope, 2012a, chapter 10).

We want to make the case that a transformation of assessment is possible if we can bring formative and summative assessments together into a single, integrated process. Pellegrino, Chudowsky and Glaser describe the process of reasoning from evidence of learner performance as an 'assessment triangle' – consisting of three interconnected points: cognition <-> observation <-> interpretation (Pellegrino et al, 2001, p. 44).

In Table II, we compare this process of evidentiary reasoning in classical testing environments with a context of assessment of scientific writing where assessment is closely integrated with performance. This broadly construed program change in assessment paradigms conceptually represents an attempt to connect assessment more closely with instruction, and is founded on progress that has been made in recent decades in our understandings of the nature of learning (Kalantzis & Cope, 2012b). Experts in a subject domain such as science organize knowledge into schemas and make sense of new information through processes of pattern recognition and argumentation about underlying processes. These are typically represented in written argumentation. Such knowledge representations are essential tools for understanding, knowledge making and knowledge communication (Greeno & Hall, 1997; Bransford et al, 2000; Pellegrino et al, 2001; Chung et al, 2006; Mislevy, 2006a; Mislevy et al, 2007). These are characteristically social processes, requiring less memory work than a capacity to work with collaborating experts, searching sources, documenting and citing material used, and distinguishing one's own reasoning from the reasoning of others (Meltzoff et al, 2009). Assessing such disciplinary practices requires new forms of assessment (Bass & Glaser, 2004; Baker, 2007).

A growing body of research also underlines the importance of formative assessment that in turn feeds directly into the learning process (Black & Wiliam, 1998; OECD Centre for Educational Research and Innovation, 2005). Research also shows that 'situated assessment' in the form of regular and multiple forms of feedback produces enhanced learning outcomes (Palinscar & Brown, 1984; Scardamalia & Bereiter, 1996; Greeno, 1998; Shepard, 2000; Windschitl, 2002).

A new wave of assessment technologies promises significant breakthroughs in the modes of assessment, and in particular the kinds of complex disciplinary performance articulated in this article. It also offers new potentials to supplement the traditional focus on summative assessment with systematic and frequent formative assessment (Cope et al, 2011).

	Classical discrete-item summative test	Coordinating formative and summative assessment of science writing
Cognition	Discrete items of knowledge, facts and concepts that a learner has committed to memory. Application of algorithmic problem solutions.	Writing as complex and synthetic knowledge process, in the case of argumentative writing, relying on available sources such as knowledge, facts and concepts a learner can readily access. Distributed knowledge is created in social relationships of complementary expertise and collaborative intelligence. Assessable scientific understanding is construed to be what the student can represent in an extended text.
Observation	The artifact of the test, a peculiar phenomenon which is different from the texts of disciplinary knowledge and the pragmatics of learning delivery. The site of observation is at the end of a learning cycle, separated from the learning in time and space.	Direct observation of disciplinary knowledge practice, in the authentic language of the discipline and using its characteristic modes of expression; multi-perspectival assessment (e.g. self, peer, teacher). Multiple assessment modes (e.g. tag, parse, map). Learning and assessment in the same time and space.
Interpretation	A highly mediated process of inference, 'knowledge' or 'understanding' that is assumed to reside in the memory of the learner.	Less mediated interpretation of disciplinary practice. No definitive cognitive conclusions beyond what can be deduced from the evidence provided in **its** representation in the form of a written or multimodal artifact. Assessable knowledge is not located in the head of the student, in what they can remember (cognitive inference), but in the documentary work of the learner as manifest in the form of a documentary scientific artifact presented in the characteristic textual forms of the discipline.

Table II. Changing assessment paradigms.

These technologies will make it possible for science students to have access to valuable formative feedback on their written scientific discourse anytime, anywhere. This is a significant advantage that provides a viable and meaningful means to bypass the feedback 'bottlenecks' associated with the centralization of assessment with a solo science teacher serving the needs of a large number of students. Scientific argument writing and peer review spaces like the one we describe later in this article help students develop a bird's-eye view of the structure of their argumentative writing, support them as they robustly elucidate the ways in which the various components of their arguments are or are not linked, and provide a means to revisit and revise both the structure and substance of their arguments. These activities, which require students to apply, analyze, synthesize, and evaluate knowledge – in short, that develop critical thinking skills – will be more durable (Knowlton, 2001). The participative nature of the argument writing and review activities – like those provided by this technology, will facilitate the construction of deep and lasting knowledge (Hacker & Niederhauser, 2000). In this project, we have explored one possibility in these new spaces of technologically mediated formative assessment, and that is web-enabled peer review as described in the remaining sections of this article.

In the near future, by distributing feedback across machine-mediated, self- and peer- and teacher-review systems, such technologies also support teachers by providing immediate and targeted formative feedback to students on written scientific argumentation at multiple stages in the drafting process. This helps teachers with the daunting and labor-intensive task of providing feedback on large numbers of student texts, thus encouraging teachers to give more assignments that involve scientific argumentation and sophisticated assessments of scientific knowledge.

Writing Science on the Web: the *Scholar* intervention

Supported by research and development grants from the Institute of Education Sciences, US Department of Education, *Scholar* is a web-writing environment which affords students the possibility to see writing rubrics, as well as make and present self, peer and teacher reviews closely integrated with their work. In the case of the research reported in this article, a scientific argument rubric based on the Common Core State Standards specification of argumentative writing for Grade 8 was created (2010).

Our theory of change in this implementation focuses on deepening student and teacher awareness and understanding of 'scientific argument'. Change processes at a micro level are illustrated in Figure 1, anticipating that change will occur in iterative cycles during which students formulate and receive formative assessment feedback on their scientific arguments, and in which teachers evaluate the effectiveness of their teaching of argument based on the arguments their students produce.

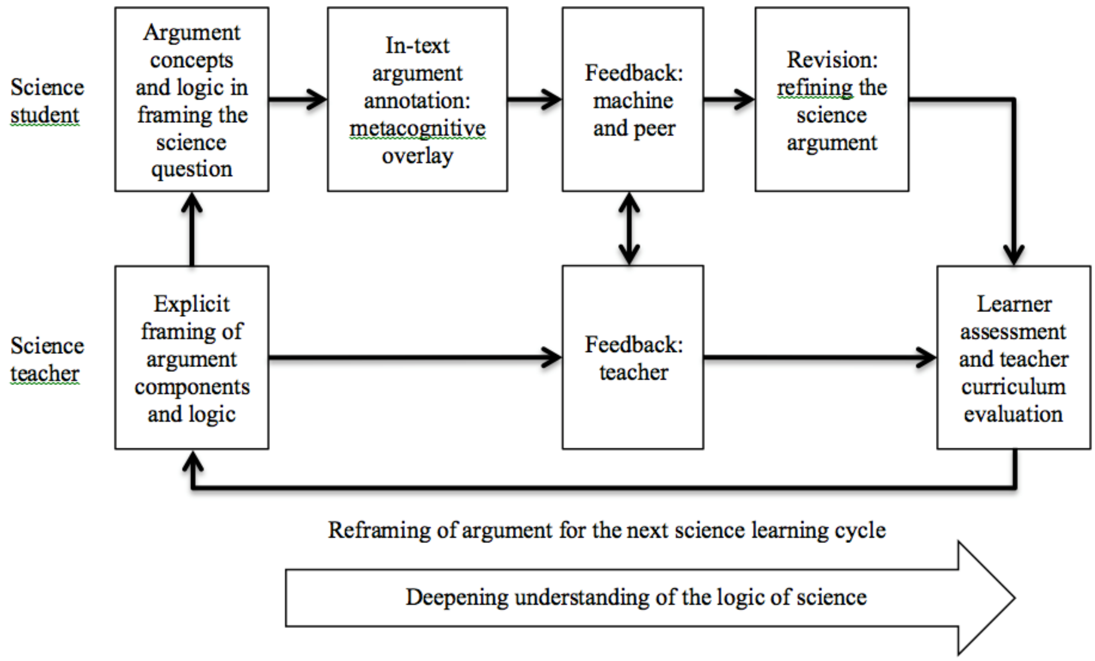

Figure 1. The microdynamics of change in writing scientific argument.

For the student, the process of change produces a deepened understanding of scientific argument, involving the following steps, lasting for the duration of a unit of work, during which they are required to represent their science knowledge as a written science argument:

1. The student is explicitly presented with key concepts of argument in the framing of the science question, and with explicit instruction about language choices with an expectation that their work will exemplify science argument.
2. At the feedback stage, students are offered formative assessment: a) machine feedback; and b) peer and teacher feedback on arguments, with a focus on offering comments and suggestions to strengthen their arguments. Peer review can be anonymous or open, depending upon the setting established by their science teacher. At this stage in the processes of students' developing and refining their argument, the *Scholar* environment is designed to offer Web 2.0 opportunities for distributed peer formative assessment, as well as prompting metacognitive reflection while analyzing the effectiveness of peer arguments.
3. Using this feedback, the student revises and refines their science argument for publication into a student- and class-level science knowledge bank.

Each time the student cycles through this process, it is anticipated that their understanding of argumentative discourse and scientific representation will deepen. For the teacher, the process of change involves evaluation of student responses to their framing of argument assignments, assessment of student arguments within the *Scholar* environment, and revision of assignments in subsequent cycles of learning.

Arguments about Fracking: a case study of students writing science

Now, we move to two Grade 8 classes in a school in New York City. Here, the research team was engaged in feasibility testing *Scholar* as a writing and formative assessment environment. This is a middle school in which 71% of students get a free lunch, 45% of children are Black, 29% are Hispanic, 20% are White, and 4% are Asian. 40% of children have Individual Education Plans.

Key research questions in our intervention were:

1. How do teachers and learners negotiate a web-based writing and peer review/formative assessment environment?
2. How might such an environment support the writing of scientific argument?

In the section that follows the teacher (the same teacher for both classes) moves through a ten-lesson research, discussion and writing cycle.

In both classes, the teacher is aiming to deepen students' and teachers' awareness and understanding of 'scientific argument'. The topic he has chosen is 'hydraulic fracking', a process of extracting natural gas which has, in recent years, produced a quite dramatic shift away from coal as an energy source in the United States, particularly for electricity generation. The environmental consequences, however, are a matter of scientific dispute.

The teacher shows the students the movie, *Gasland* (http://www.gaslandthemovie.com), exploring the social and environmental consequences of fracking. He asks students to research the science and surrounding arguments about fracking, ranging from community groups, to the petrochemical industry. Over several weeks, the students:

1. Research scientific claims, evidence and alternative arguments.
2. Write a draft of their argument, during which they are exposed to review criteria and rating scales based on the Common Core Standards for argument.
3. Peer review – assess each other's work, offering quantitative ratings and qualitative commentary.
4. Rewrite, based on the reviews received from peers and the experience of reviewing other students' texts.

In the case we describe here, 30 'fracking' texts were written in *Scholar*, across two classes. While writing their first drafts, students were exposed to review criteria for writing scientific argument created by the teacher and researchers on the right side of the screen, as represented in Figure 2. The teacher and researcher opted for a four-step rating scale, 0 to 3 across six evenly weighted review criteria: Introduction, Claims, Evidence, Counter-claims, Conclusions and Conventions (see Table III). To guide the reviewers and to optimize inter-rater reliability, they also provided the *Scholar* review criteria as pull-down information menus, rating descriptions for each of the four rating levels for each criterion.

Next, students accessed each other's works anonymously, randomly assigned by *Scholar*, and wrote reviews of two others students' works, dragging the slider to give a rating, providing feedback on each criterion and offering an explanation of the rating they had given, as represented in Figure 3. Following this, the students received reviews from the two other students, as well as an aggregated score for each criterion across two reviewers, as represented in Figure 4.

With the help of this feedback, students then rewrote the text, ready for final review by the teacher and publication by the teacher into the student's web-writing portfolio.

The pedagogical architecture of *Scholar* reflects a series of left/right juxtapositions within the screen: L: the work; R: social and evaluative conversations around the work; L: cognition; R: metacognition; L: individual work; R: social dialogue which assists in the collaborative (re)design of works. As they wrote, the students had direct access to the review criteria presented in Table III. They then reviewed two works by their peers according to these criteria.

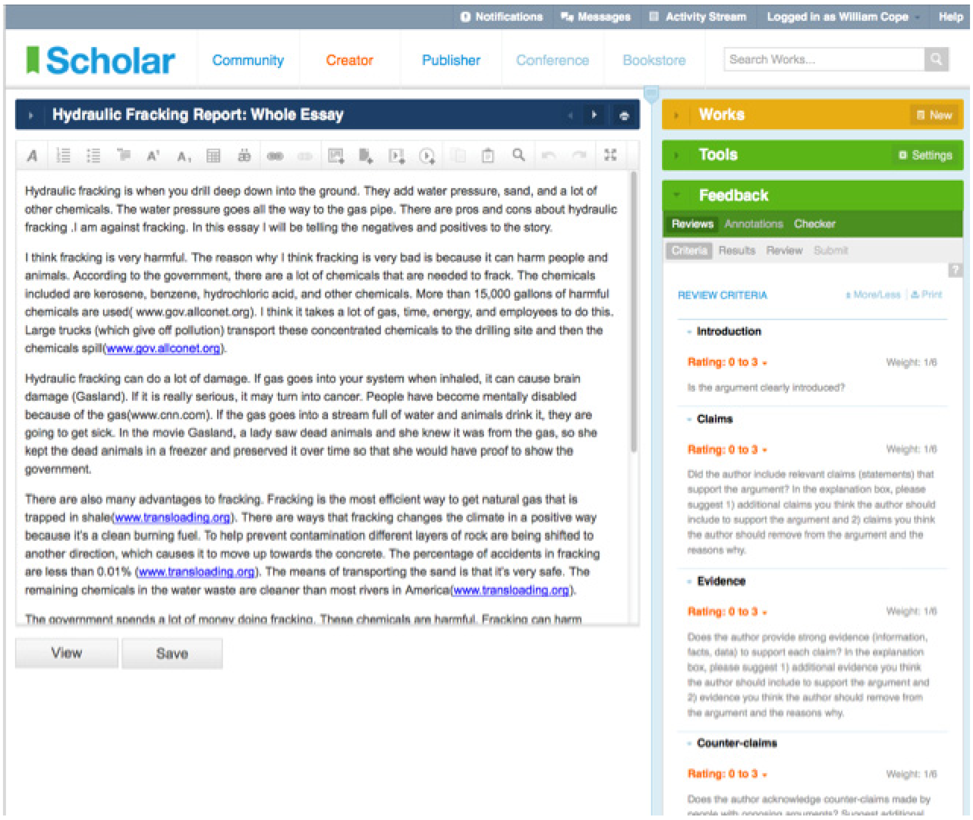

Figure 2. First draft text: work on the left; teacher-constructed review criteria juxtaposed on the right.

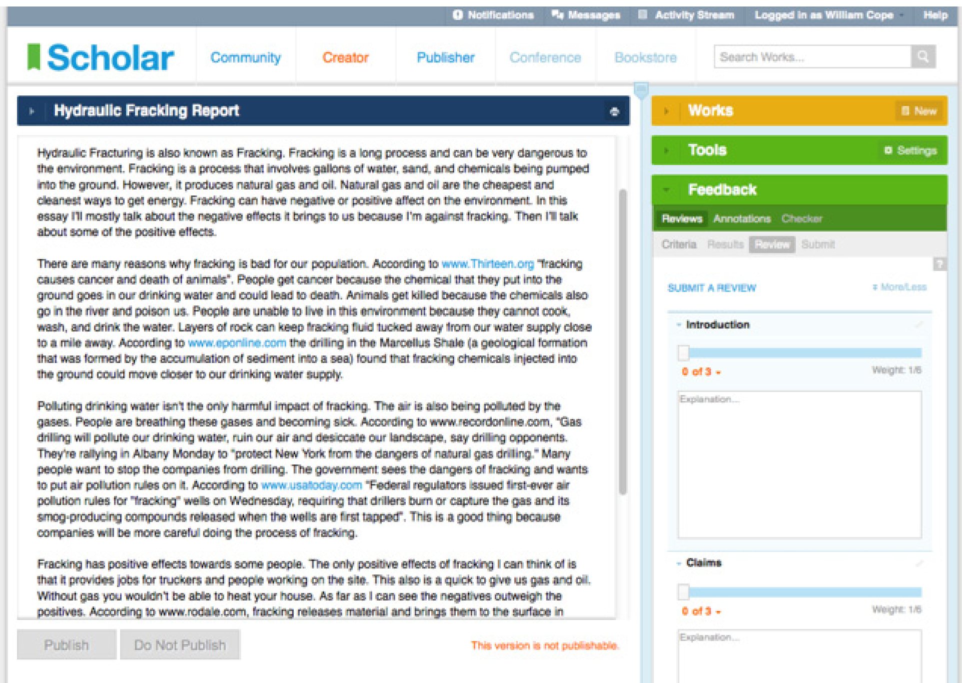

Figure 3. Reviewing a peer's text.

Figure 4. Receiving review feedback.

Criterion	Description	Rating descriptions
Introduction	Is the argument clearly introduced?	0. Argument is not introduced 1. Argument is poorly introduced. 2. Argument is introduced, but not clearly. 3. Argument is clearly introduced.
Claims	Did the author include relevant claims (statements) that support the argument? In the explanation box, please suggest: 1. additional claims you think the author should include to support the argument; and 2. claims you think the author should remove from the argument and the reasons why.	0. No claims are included. 1. A claim is included but it does not support the argument. 2. Claims are included and they mostly support the argument. 3. A sufficient number of relevant claims is included and they all support the argument.
Evidence	Does the author provide strong evidence (information, facts, data) to support each	0. Provides no evidence to support each claim. 1. Presents some evidence, but not all evidence

	claim? In the explanation box, please suggest: 1. additional evidence you think the author should include to support the argument; and 2. evidence you think the author should remove from the argument and the reasons why.	supports each claim. 2. Presents some evidence to support each claim, but the argument would be stronger with additional evidence. 3. Presents strong evidence to support each claim.
Counter-claims	Does the author acknowledge counter-claims made by people with opposing arguments? Suggest additional counter-claims that should be acknowledged and refuted.	0. Counter-claims not acknowledged. 1. Some counter-claims acknowledged but not refuted. 2. Some relevant counter-claims acknowledged and refuted, but the opportunity to refute some important counter-arguments has been missed. 3. Main counter-claims and opposing arguments thoroughly refuted.
Conclusions	How well does the author tie the argument together? Suggest possible changes and improvements.	0. There is no conclusion or it is unrelated to the argument. 1. Conclusion is difficult to follow or unrelated to the argument. 2. Conclusion is clearly worded, but may deal with more than one subject or have emotional or prejudicial phrases. 3. Conclusion is clearly worded, deals with one subject, and is free of emotional or prejudicial phrases.
Conventions	Did the author use appropriate grammar, spelling, punctuation, paragraphing, and capitalization? Use the annotations tool to suggest changes.	0. Errors prevent reader from following argument. 1. Errors make the argument unclear. 2. Few errors. 3. Almost no errors.

Table III. Review criteria.

Here we analyze quantitative ratings data and qualitative textual data from this intervention, writing science arguments in a web environment that enables peer review. First, the overall ratings data. For us, the most interesting questions are the amount of inter-rater disagreement and the amount of disagreement between peers and expert raters from the research team. These data are presented in Table IV.

	Introduction	Claims	Evidence	Counter-claims	Conclusion	Conventions	Overall mean
Mean peer disagreement	.38	.54	.77	.85	.42	.69	.60
Mean disagreement of expert rater with the mean peer rating	.21	.33	.30	.62	.62	.74	.47

Table IV. Peer disagreement (mean distance between peers' scores), and disagreement of expert raters with peers (mean distance between the expert rater's score and the mean score of peers).

On a four-level rating scale, the mean peer disagreement never exceeds 0.85 rating levels, and on one criterion is as low as 0.38. The overall mean disagreement is 0.6 of a rating level – which, with a variation representing just 15% of the scale, must be considered quite a low level of disagreement.

Even more interesting is the amount of disagreement between an expert rater who graded the texts and the mean rating of the reviewers. Here, the disagreement never exceeds 0.74 of a rating level, and is as low as 0.21 of a rating level. The disagreement between the expert rater and the mean score of the two student raters is only 0.47 of a rating level, a variation representing 12%

of the scale. In other words, within a relatively small margin of error, peer ratings might be taken to be a proxy for expert ratings. The reliability of the peer ratings might be taken to be a reflection of the fact that the review criteria and rating levels were spelt out explicitly in *Scholar,* and that students had already become familiar with these as they were exposed to them during the writing of their own texts.

Now to analyze discourse patterns in the texts themselves, criterion by criterion. The texts extracted below are representative of the range of textual moves that students in the sample classes were making as they made their arguments.

An argument text typically starts with an introduction, in which the writer poses the question or problem that their argument sets out to address, outlines their argument, and states their thesis. This was the area with the highest level of peer agreement on rating level, and then highest level of peer and expert rater agreement. Here is one of the higher rated texts:

> Hydraulic Fracturing is also known as Fracking. Fracking is a long process and can be very dangerous to the environment. Fracking is a process that involves gallons of water, sand, and chemicals being pumped into the ground. However, it produces natural gas and oil. Natural gas and oil are the cheapest and cleanest ways to get energy. Fracking can have negative or positive affect on the environment. In this essay I'll mostly talk about the negative effects it brings to us because I'm against fracking. Then I'll talk about some of the positive effects. (Student A6)

This paragraph briefly describes the process, links the shorter and more colloquial term to the full scientific term, and introduces the subsequent discussion about environmental debate and foreshadows the author's own judgment.

Argument texts support their theses with claims, or supporting statements that address the argument. This was an area in which there was more agreement amongst peer raters, and between peer raters and the expert rater. For instance:

> Hydraulic fracking can do a lot of damage. (Student A1)
>
> [First claim] Fracking is bad for many reasons involving humans.
>
> [Second claim] Fracking is bad for animals for numerous reasons. (Student B12)

Claims need to be backed up with evidence, using and citing credible sources and providing appropriate scientific data to support each claim. This was another area in which there was less disagreement amongst peer raters than some other criteria, and between peer raters and the expert rater. Some examples:

> [Claim:] The second reason im against fracking is becuase of the harmful affects it has on humans, and animals. [Supporting evidence:] Becuase of the chemicals that are in fracking like: Crystalline Silicia, and Diesel can cause people to get cancer. Some of these chemicals cause prblems in the reproductive system as well.(active post.com). There are other chemicals like: Methanol, Naphthalene, Formaldehyde, Lead and Sulfuric acid, can cause other serious illnesses or deaths. Scientist beleive that Fracking is causing animals to die (that drink water). A professor and his team examined 24 cases of sickened animals in 6 states. They found reproductive problems in birds, cats, chickens, cows, deers, horses, koi, and llamas.(Cornell U). A women said that she's found dead animals and has been keeping them in her freezer until they can get an autopsy on them to see the cause of their death.(Gasland). So thats how Fracking affects animals and humans. (Student B14)

Claims are tied to data by chains of evidentiary *reasoning* or warrants, not explicitly measured in the review criteria used in this class. Here claims are supported with logical, scientific reasoning, such as a justification that connects the evidence to the claim and shows why the data counts as evidence by applying scientific principles. For instance:

> [Claim:] Fracking is bad for your health because it causes diseases. [Chain of reasoning from evidence:] It causes cancer and sometimes it gives children asthma. Fracking causes these diseases because during the fracking process it goes through the ground and into people's water well's. When the water goes through the well's the air gets contaminated with the fumes of the fracking gas. So when that happens people start breathing in poison air, and when they breathe it in that's what causes cancer and asthma. The fracking gas is bad for the environment because

animals breathe the same air we breathe. So when they breathe the air they get severe diseases. Then over time the animals die. Isn't that sad the animals suffer for years until they die. (GasLand). (Student A11)

Effective scientific arguments acknowledge counter-claims from alternate or opposing positions, distinguish these claims from those being made in the current argument, and rebut those claims that are judged to be questionable. There was somewhat less agreement amongst peer raters and expert raters about the effectiveness of presentation of counter-claims. An example of a counter-claim:

[Counter-claim:] There are many positive about fracking but it doesn't out way the negatives of fracking. [Evidence supporting the counter-claim:] According to many companies participate in the search and recovery of deep shale liquids (oil, natural gas, unconventional fuels). (Blogspot.com) it's a mining or drilling technique used to break up rock underground to create easier access to resources. (heatingoil.com) it is even used sometimes to revive flagging drinking water wells. (heatingoil.com) it's also used for some solid (as opposed to liquid or gas) mineral resources. (heatingoil.com) Fracking will also lead to natural gas being substituted for coal, which is a net win on all sorts of pollutants, including carbon dioxide. (Student B8)

A scientific argument is tied together by a conclusion in the form of an overall evaluative judgment. Peers agreed more about the ratings of conclusions than expert raters with peers. Here's a conclusion that attempts scientific balance and recommends a way forward, consistent with the rubric or 'review criteria' presented to the students:

Now that we found a cheaper and quicker way to get natural gas is all of those chemicals that are hurting the animals and humans that are breathing it in who get permit brain damage or cancer is all of that damage is worth it? Even though we do depend on natural gas since oil is running out they should find a better way to use fracking instead of using chemicals that harm people all over the world. Since fracking saved us money and saved us oil we can do more things with natural gas now. Even though there is a lot of a bad and good thing about fracking a lot of things that need to change about fracking. The one thinghat they need to change is fracking needs to change is the chemicals. Another thing that they should change about fracking is that they are putting fracking near people house and when they do that they are affecting people water and the air that they breathe. Hopefully fracking changes over time but we do need fracking even though it is hurting the environment at the same time. [sic] (Student A9)

The teacher also asked students to review each other's work for writing conventions. The Common Core State Standards define these, in the case of written argument as 'a formal style' consisting of 'words, phrases, and clauses to create cohesion and clarify the relationships among claim(s), counter-claims, reasons, and evidence'. The conventions of argument involve the kinds of depersonalized discourse described earlier in this article in the reference to Gee (2004a), the forms of abstraction described by Lemke (1990) and the 'deliberate semantics' described by Vygotsky (1986). Perhaps reflecting a greater difficulty to detect these conventional forms, this was the area of rating with the greatest peer disagreement and expert disagreement with peers. Also, a general observation might be made: that on the measure of writing conventions, these texts are far less successful than they are in their mastery and analysis of the logic of written scientific argument. Following are some more successful examples of characteristically scientific writing, the features of which we described in the earlier sections of this article and which fall under the CCSS notion of 'formal writing':

Natural Gas is one of the principle sources of energy for many of our day-to-day needs and activities. Hydraulic fracturing is the propagation of fractures in a rock layer caused by the presence of a pressurized fluid. It is also known as fracking. (Student B8)

The whole process of fracking is extremely dangerous, and harmful to the environment. ... As you can see, fracking doesn't just affect humans, animals also. (Student B7)

Even though it [fracking] has a lot of negatives it also has positive effects on the environment and government. According to aboutnaturalgas.com 'natural gas is providing the united states with an enormous economic advantage as a result of American ingenuity. (Student A5)

> Natural Gas is effecting the health of people, and it is even effecting the lives of them as well. ... Fracking is now becoming a huge threat to our environment. In the movie, Gasland, a lot of animals were losing hair because of dehydration that resulted in the water turning into natural gas. (Student A10)

> Fracking for natural gas is killing domestic animals like horses, cattle, goats, sheep. The dead animals provide a strong warning that fracking can harm humans (wayneglessner.com). (Student B10)

> And here, on the same measure, are some less successful examples, in these cases reflecting aspects of 'informality':

> Fracking is not something to play around with if you don't know what to do. ... We shold stop now or eles. (Student A2)

> The only positive effects of fracking I can think of is that it provides jobs for truckers and people working on the site. ... They only care about the money so they pay them to not say anything about fracking so the company won;t look good, but fracking is messed up and they should fix what they are doing wrong. (Student A7)

> Personally I don't agree with Fracking. In this essay I would tell you why I don't agree with it. (Student A3)

> Let me tell you my opinion on fracking. Fracking to me is bad for the environment and good for the economy like I said. It is good for the economy because it gives people more jobs in New York, and gives more money to the government. Fracking is bad for the planet because it kills helpless creatures. It kills helpless creatures by the poison gas in the air. So you see why I'm struggling with this. I think fracking is good and bad at the same time. (Student A11) [Student Reviewer 1 response:] You should stop saying that u are going to explain and just do it. ... Don't say u are going to tell us. just say it, say how its bad and just say that.

One of the characteristic features of the *Scholar* environment is the juxtaposed metarepresentations, or representations about representation. These take the form of the review criteria that sit beside the text under construction, then the peer review process itself, followed by revision in which the students take these reviews into account. As discussed earlier in this article, metarepresentation is a tool for uncovering the principles of order in texts, in this case how scientific arguments work, who they work for, and why they work for them.

If the texts these students created were weak in terms of the narrowly defined writing conventions of spelling, grammar and punctuation, the students mostly showed a keenly perceptive understanding of the nature of argumentation, even in a field of unfamiliar scientific knowledge:

> I like the intro, but I think you could have had done better with the reason of why you chose to write about this essay. (Reviewer 2, Student Text A7)

> The introduction is very good. It tells the thesis and explains whats going to be in the essay. The intro tells whats going to be in the essay, but it doesnt tell the whole essay. (Reviewer 1, Student Text A12)

> the author needs to back up there statement. like when they said 'fracking is the easiest way to serve energy to people in the world, and the cleanest way to get energy'. (Reviewer 1, Student Text A2)

> I feel this person did very good with evidence because they said how the felt about fracking and then proved it by displaying facts on their topic and then backed some of this persons facts with quotes and who quoted them. (Reviewer 1, Student Text B7)

> When grading this i didnt see alot of evidence. I didnt see the evidence because the person only put cnn.com. I didnt see any quotes and alot of information. The person didnt put a work cited page. He didnt put the page that tells all the websites he got the information from. You cant tell if he cant back it up resonably. (Reviewer 2, Student Text A15)

I feel this person did very good with evidence because they said how the felt about fracking and then proved it bydisplaying facts on their topic and then backed some of this persons facts with quotes and who quoted them. This person did very well but I wish they would have shown more of another sides argument because I feel this person did more of what they feel and then just backed it up with some quotes. (Reviewer 1, Student Text B7)

Conclusion is clearly worded, deals with one subject, and is free of emotional or prejudicial phrases. (Reviewer 1, Student Text A7)

Conclusions

In this article, we have analyzed the process of writing scientific argument as an instance of complex epistemic performance, as well as the dynamics of formative and summative assessment of this performance. We go on to describe a classroom in which students are writing scientific arguments and providing each other with peer formative assessment, during a process of drafting, review and revision.

Several principal conclusions emerge. One is that there is a strikingly low level of disagreement on rating of assessment criteria between peers and even less between peers and an expert assessor. This indicates that peer review may have a valid place in new assessment regimes.

The second conclusion is that, even when assessment of writing, based on narrowly conceived conventions such as spelling, grammar and punctuation may produce less than happy results for some students, these same students may nevertheless develop impressive capacities for written scientific argumentation. In this case, the conditions for the development of these capacities were exposure to a pedagogical frame offering explicit review criteria during the first draft, reviewing another student's work against these criteria, then rewriting in order to strengthen the texts with a clearer understanding of the essence of these criteria. The keys to success here were metacognitive awareness, formative assessment, and the collaborative process of giving and receiving feedback. In these conditions, one might argue that the students have acquired, despite their relative weakness on conventional measures of writing, a clear sense of one of the disciplinary foundations of science – scientific argumentation.

Notes

[1] We wish to acknowledge funding support from the US Department of Education Institute of Education Sciences: 'The Assess-as-You-Go Writing Assistant: a student work environment that brings together formative and summative assessment' (R305A090394); 'Assessing Complex Performance: a postdoctoral training program researching students writing and assessment in digital workspaces' (R305B110008); 'u-learn.net: an anywhere/anytime formative assessment and learning feedback environment' (ED-IES-10-C-0018); 'The Learning Element: a lesson planning and curriculum documentation tool for teachers' (ED-IES-10-C-0021); and 'Infowriter: a student feedback and formative assessment environment for writing information and explanatory texts' (ED-ED-IES-13-C-0039). We also wish to acknowledge funding support from the Bill and Melinda Gates Foundation. *Scholar* code owned by the University of Illinois has been licensed by Common Ground Publishing LLC, directed by Bill Cope and located in the Research Park at the University of Illinois.

References

Abell, Michael (2006) Individualizing Learning Using Intelligent Technology and Universally Designed Curriculum, *Journal of Technology, Learning, and Assessment*, 5.

Abell, S., Anderson, G. & Chezem, J. (2000) Science as Argument and Explanation: exploring concepts of sound in third grade, pp. 100-119, in J. Minstrell & E.H.V. Zee (Eds) *Inquiry into Inquiry Learning and Teaching in Science*. Washington: AAAS.

Abi-El-Mona, Issam & Abd-El-Khalick, Fouad (2010) Perceptions of the Nature and 'Goodness' of Argument among College Students, Science Teachers, and Scientists, *International Journal of Science Education*, iFirst Article, 1-33.

American Association for the Advancement of Science (1993) *Benchmarks for Science Literacy*. New York: Oxford University Press.

Baker, Eva L. (2004) *Aligning Curriculum, Standards, and Assessments: fulfilling the promise of school reform*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.

Baker, Eva L. (2007) Moving to the Next Generation System Design: integrating cognition, assessment, and learning. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.

Bass, Kristin M. & Glaser, Robert (2004) *Developing Assessments to Inform Teaching and Learning*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.

Bell, P. & Linn, M. (2000) Scientific Arguments as Learning Artifacts: designing for learning from the web with KIE, *International Journal of Science Education*, 22, 797-817. http://dx.doi.org/10.1080/095006900412284

Black, P. & Wiliam, D. (1998) Assessment and Classroom Learning, *Assessment in Education*, 5, 7-74. http://dx.doi.org/10.1080/0969595980050102

Bowker, Geoffrey C. & Star, Susan Leigh (2000) *Sorting Things Out: classification and its consequences*. Cambridge MA: MIT Press.

Bransford, John D., Brown, Ann L. & Cocking, Rodney R. (2000) How People Learn: brain, mind, experience and school. N.R.C. Commission on Behavioral and Social Sciences and Education. Washington: National Academy Press.

Brown, J., Collins, A. & Duguid, P. (1989) Situated Cognition and the Culture of Learning, *Educational Researcher*, 18.

Cazden, Courtney B. (2001) *Classroom Discourse: the language of teaching and learning*. Portsmouth, NH: Heinemann.

Cetina, Karin Knorr (1999) *Epistemic Cultures: how the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Chung, Gregory K.W.K., Baker, Eva L., Brill, David G., Sinha, Ravi, Saadat, Farzad & Bewley, William L. (2006) Automated Assessment of Domain Knowledge with Online Knowledge Mapping. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.

Common Core State Standards Initiative (2010) Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington DC. http://www.corestandards.org/

Cope, Bill, Kalantzis, Mary, McCarthey, Sarah, Vojak, Colleen & Kline, Sonia (2011) Technology-mediated Writing Assessments: paradigms and principles, *Computers and Composition*, 28, 79-96. http://dx.doi.org/10.1016/j.compcom.2011.04.007

Darling-Hammond, Linda, Ancess, Jacqueline & Falk, Beverly (1995) *Authentic Assessment in Action: studies of schools and students at work*. New York: Teachers College Press.

Darling-Hammond, Linda & Wood, George H. (2008) *Assessment for the 21st Century: using performance assessments to measure student learning more effectively*. Stewart, OH: Forum for Education and Democracy.

Driver, Rosalind, Newton, Paul & Osborne, Jonathan (2000) Establishing the Norms of Scientific Argumentation in Classrooms, *Science Education*, 84, 287-312. http://dx.doi.org/10.1002/(SICI)1098-237X(200005)84:3%3C287::AID-SCE1%3E3.0.CO;2-A

Duschl, R. (1998) Research on the History and Philosophy of Science, in B.J. Fraser & K.G. Tobin (Eds) *International Handbook of Science Education*, pp. 443-465. Dordrecht: Kluwer.

Duschl, R. & Ellenbogen, K. (2001) Scaffolding and Assessing Argumentation Processes in Science Learning. Paper presented at the Bi-annual Meeting of the European Association for Research on Learning and Instruction, Friborg, September.

Duschl, R. & Osborne, J. (2002) Supporting and Promoting Argumentation Discourse in Science Education, *Studies in Science Education*, 38, 39-72. http://dx.doi.org/10.1080/03057260208560187

Eurduran, S. & Jimenez-Aleixandre, M.P. (2008) Argumentation in Science Education: an overview, in S. Eurduran & M.P. Jimenez-Aleixandre (Eds) *Argumentation in Science Education*, pp. 3-28. New York: Springer.

Etkina, Eugenia, Mestre, Jose P. & O'Donnell, Angela (2005) The Impact of the Cognitive Revolution on Science Learning and Teaching, in J. Royer (Ed.) *The Impact of the Cognitive Revolution on Educational Psychology*, pp. 119-164. Greenwich, CT: Information Age Publishing.

Gee, James Paul (1992) *The Social Mind: language, ideology, and social practice*. New York: Bergin & Garvey.

Gee, James Paul (2004a) Language in the Science Classroom: academic social languages as the heart of school-based literacy, in E.W. Saul (Ed.) *Crossing Borders in Literacy and Science Instruction*, pp. 13-32. Arlington, VA: NSTA Press.

Gee, James Paul (2004b) *Situated Language and Learning: a critique of traditional schooling*. London: Routledge.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D. & Brenwald, S. (2008) Highlights from TIMSS 2007: mathematics and science achievement of US fourth- and eighth-Grade students in an international context. Washington: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

Greeno, James G. (1998) The Situativity of Knowing, Learning, and Research, *American Psychologist*, 53, 5-26. http://dx.doi.org/10.1037/0003-066X.53.1.5

Greeno, James G. & Hall, Rogers P. (1997) Practicing Representation: learning with and about representation, *Phi Delta Kappan*, 78, 361-367.

Grotzer, Tina (2009) Learning to Reason about Evidence and Explanations: promising directions in education, in E. Callan, T. Grotzer, J. Kagan, R.E. Nisbett, D.N. Perkins & L.S. Shulman, *Education and a Civil Society: teaching evidence-based decision making*, pp. 51-74. Washington: American Academy of Arts and Sciences.

Hacker, D. & Niederhauser, D. (2000) Promoting Deep and Durable Learning in the Online Classroom, *New Directions for Teaching and Learning*, 84, 53-63. http://dx.doi.org/10.1002/tl.848

Halliday, M.A.K. (2004) The Language of Science, in J.J. Webster (Ed.) *The Collected Works of M.A.K. Halliday*. London: Continuum.

Halliday, M.A.K. & Martin, J.R. (1993) *Writing Science: literacy and discursive power*. London: Falmer Press.

Hogan, K. & Fisherkeller, J. (2005) Dialogue as Data: assessing students' scientific reasoning with interactive protocols, in J.J. Mintzes, J.H.Wandersee & J.D. Novak (Eds) *Assessing Science Understanding: a human constructivist view*. Burlington, MA: Elsevier Academic Press. http://dx.doi.org/10.1016/B978-012498365-6/50007-X

Jimenez-Aleixandre, Maria Pilar, Diaz de Bustamante, J. & Duschl, R. (1998) Scientific Culture and School Culture: epistemic and procedural components. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, San Diego, CA, March.

Jimenez-Aleixandre, Maria Pilar & Eurduran, Sibel (2008) Argumentation in Science Education: an overview, in S. Eurduran & M.P. Jimenez-Aleixandre (Eds) *Argumentation in Science Education: perspectives from classroom-based research*. Berlin: Springer.

Jonassen, David H. & Kim, Bosung (2009) Arguing to Learn and Learning to Argue: design justifications and guidelines, *Educational Technology Research and Development*, 58, 439-457. http://dx.doi.org/10.1007/s11423-009-9143-8

Kalantzis, Mary & Cope, Bill (2012a) *Literacies*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139196581

Kalantzis, Mary & Cope, Bill (2012b) *New Learning: elements of a science of education*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139248532

Knowlton, D. (2001) Promoting Durable Knowledge Construction through Online Discussion, in *Proceedings of the Annual Mid-South Instructional Technology Conference,* Murfreesboro TN, April, pp. 1-13.

Kolsto, S. & Ratcliffe, M. (2008) Social Aspects of Argumentation, in S. Eurduran & M.P. Jimenez-Aleixandre (Eds) *Argumentation in Science Education*, pp. 91-116. New York: Springer.

Krajcik, Joseph S. & Sutherland, LeeAnn M. (2010) Supporting Students in Developing Literacy in Science, *Science*, 328, 456-459. http://dx.doi.org/10.1126/science.1182593

Kuhn, D. (1991) *The Skills of Argument*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571350

Kuhn, D. (1993) Science as Argument: implications for teaching and learning scientific thinking, *Science Education*, 77, 257-272.

Kuhn, D. (2004) What is Scientific Thinking and How Does it Develop?, in U. Goswami (Ed.) *Blackwell Handbook of Childhood Cognitive Development,* pp. 371-393. Malden, MA: Blackwell.

Kuhn, Deanna, Goh, Wendy, Iordanou, Kalypso & Shaenfield. David (2008) Arguing on the Computer: a microgenetic study of developing argument skills in a computer-supported environment, *Child Development*, 79, 1310-1328. http://dx.doi.org/10.1111/j.1467-8624.2008.01190.x

Latour, Bruno & Woolgar, Steve (1986) *Laboratory Life: the construction of scientific facts*. Princeton NJ: Princeton University Press.

Lave, Jean & Wenger, Etienne (1991) *Situated Learning: legitimate peripheral participation*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815355

Lemke, Jay L. (1990) *Talking Science: language, learning and values*. Westport, CN: Ablex.

Lemke, Jay L. (2004) The Literacies of Science, in E.W. Saul (Ed.) *Crossing Borders in Literacy and Science Instruction*, pp. 33-47. Arlington VA: NSTA Press.

Mason, L. & Santi, M. (1994) Argumentation Structure and Metacognition in Shared Knowledge at School. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April.

Meltzoff, Andrew N., Kuhl, Patricia K., Movellan, Javier & Sejnowski, Terrence J. (2009) Foundations for a New Science of Learning, *Science*, 325, 284-288. http://dx.doi.org/10.1126/science.1175626

Mestre, Jose P., Ross, Brian H., Brookes, David T., Smith, Adam D. & Nokes, Timothy (2009) How Cognitive Science Can Promote Conceptual Understanding in Physics Classrooms, in I.M. Saleh & M.S. Khine (Eds) *Fostering Scientific Habits of Mind: pedagogical knowledge and best practices in science education*, pp. 145-171. Rotterdam: Sense.

Mislevy, Robert J. (2006a) Cognitive Psychology and Educational Assessment, in R.L. Brennan (Ed.) *Educational Measurement*, pp. 257-305. New York: Praeger.

Mislevy, Robert J. (2006b) *Issues of Structure and Issues of Scale in Assessment from a Situative/Sociocultural Perspective*. Los Angeles: Center for the Study of Evaluation National Center for Research on Evaluation, University of California.

Mislevy, Robert J., Behrens, John T., Bennett, Randy E., Demark, Sarah F., Frezzo, Dennis C., Levy, Roy et al (2007) *On the Roles of External Knowledge Representations in Assessment Design*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.

Munford, D. & Zembal-Saul, C. (2002) Learning Science through Argumentation: prospective teacher's experiences in an innovative science course. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, New Orleans, LA, April.

National Research Council (2011) *A Framework for K-12 Science Education: practices, crosscutting concepts, and core ideas*. Washington: Committee on a Conceptual Framework for New K-12 Science Education Standards, Board on Science Education, Division of Behavioral and Social Sciences and Education.

National Science Board (2010) *Science and Engineering Indicators 2010*. Arlington, VA: National Science Foundation.

Newell, George E., Beach, Richard, Smith, Jamie & VanDerHeide, Jennifer (2011) Teaching and Learning Argumentative Reading and Writing: a review of research, *Reading Research Quarterly*, 46, 273-304.

Next Generation Science Standards (2013) Washington DC: The National Academies Press. http://nextgenscience.org/next-generation-science-standards

OECD (2010) *PISA 2009: program for international student assessment results*. Paris: OECD Centre for Educational Research and Innovation.

OECD Centre for Educational Research and Innovation (2005) *Formative Assessment: improving learning in secondary classrooms*. Paris: Organisation for Economic Co-operation and Development.

Osborne, Jonathan (2005) The Role of Argumentation in Science Education, in K. Boersma (Ed.) *Research and the Quality of Science Education*, pp. 367-380. Springer: Amsterdam.

Palinscar, Annemarie Sullivan & Brown, Ann L. (1984) Reciprocal Teaching of Comprehension-fostering and Comprehension-monitoring Activities, *Cognition and Instruction*, 1, 117-175. http://dx.doi.org/10.1207/s1532690xci0102_1

Patronis, T. & Spiliotopoulou, V. (1999) Students' Argumentation in Decision Making on a Socio-Scientific Issue: implications for teaching, *International Journal of Science Education*, 21, 745-754. http://dx.doi.org/10.1080/095006999290408

Pellegrino, James W., Chudowsky, Naomi & Glaser, Robert (2001) Knowing What Students Know: the science and design of educational assessment. Washington: National Academies Press.

Pontecorvo, C. (1987) Discussing for Reasoning: the role of argument in knowledge construction, in E.D. Corte, H. Lodew, R. Parmentier & P. Span, *Learning and Instruction: European research in an international context*, pp. 239-250. Oxford: Pergamon.

Scardamalia, Marlene & Bereiter, Carl (1996) Student Communities for the Advancement of Knowledge, *Communications of the ACM*, 39, 36-37. http://dx.doi.org/10.1145/227210.227220

Shepard, Lorrie A. (2000) *The Role of Classroom Assessment in Teaching and Learning*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California.

Toulmin, Stephen (2003) *The Uses of Argument*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511840005

Van Eemeren, Frans H. & Grootendorst, Rob (2004) *A Systematic Theory of Argumentation: the pragma-dialectical approach*. Cambridge: Cambridge University Press.

Vygotsky, L.S. (1978) *Mind in Society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press. (Original work published 1962.)

Vygotsky, Lev (1986) *Thought and Language*. Cambridge, MA: MIT Press. (Original work published 1934.)

Wagner, Tony (2008) *The Global Achievement Gap: why even our best schools don't teach the new survival skills our children need, and what we can do about it*. New York: Basic Books.

Wenger, Etienne, McDermott, Richard & Snyder, William M. (2002) *Cultivating Communities of Practice: a guide to managing knowledge*. Cambridge MA: Harvard Business School Press.

Windschitl, Mark (2002) Framing Constructivism in Practice as the Negotiation of Dilemmas: an analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers, *Review of Educational Research*, 72, 131-175. http://dx.doi.org/10.3102/00346543072002131

Zeidler, D. (1997) The Central Role of Fallacious Thinking in Science Education, *Science Education*, 81, 483-496. http://dx.doi.org/10.1002/(SICI)1098-237X(199707)81:4%3C483::AID-SCE7%3E3.0.CO;2-8

Zembal-Saul, C. (2005) Pre-service Teachers' Understanding of Teaching Elementary School Science as Argument. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Dallas, TX, April 4-7.

Zembal-Saul, C. & Land, S. (2002) Scaffolding the Construction of Scientific Arguments by Preservice Teachers in a Problem-based Environment. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April.

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. http://dx.doi.org/10.1002/tea.10008

BILL COPE is a professor in the Department of Educational Policy Studies at the University of Illinois. He is also Director of Common Ground Publishing, located in the Research Park at the University of Illinois, developing the internet publishing software *Scholar* for schools and scholarly publications. Recent books include *The Future of the Academic Journal,* edited with Angus Phillips (Chandos, 2009) and *Towards a Semantic Web: connecting knowledge in academic research*, co-authored with Kalantzis and Magee (Woodhead, 2010). With Mary Kalantzis, he is co-author or editor of: *Multiliteracies: literacy learning and the design of social futures* (Routledge, 2000); *New Learning: elements of a science of education* (Cambridge University Press, 2008/2012); *Ubiquitous Learning* (University of Illinois Press, 2009); and *Literacies* (Cambridge University Press, 2012).
Correspondence: bill.cope@illinois.edu

MARY KALANTZIS is Dean of the College of Education at the University of Illinois at Urbana-Champaign. She was formerly Dean of the Faculty of Education, Language and Community Services at RMIT University in Melbourne, Australia, and President of the Australian Council of Deans of Education. With Bill Cope, she is co-author or editor of: *Multiliteracies: literacy learning and the design of social futures* (Routledge, 2000); *New Learning: elements of a science of education* (Cambridge University Press, 2008/2012); *Ubiquitous Learning* (University of Illinois Press, 2009); and *Literacies* (Cambridge University Press, 2012).

FOUAD ABD-EL-KHALICK is Professor of Science Education and Head of the Department of Curriculum and Instruction, College of Education, University of Illinois at Urbana-Champaign. His research focuses on teaching and learning about the nature of science (NOS) in precollege grades, and in preservice and inservice science teacher education settings.

ELIZABETH BAGLEY earned joint PhDs in Educational Psychology and Environment and Resources at the University of Wisconsin-Madison. As a post-doctoral researcher, she collaborated with Dr Cope on his Assess-as-You-Go project. Dr Bagley is currently a learning designer at LeapFrog Enterprises. In her role at LeapFrog, she uses her science education expertise to design and develop games, books and DVDs that help children reach their potential.