Chapter 8
Argumentation and Evaluation Intervention in Science Classes: Teaching and Learning with Toulmin

Janis A. Bulgren and James D. Ellis

Introduction

A major challenge for teachers in our schools is to help students engage in scientific reasoning (National Research Council (NRC), 1996; National Research Council (NRC), 2007). One aspect of scientific reasoning is the ability to evaluate claims or statements made about scientific issues in a variety of fields. To evaluate claims, students must use reasoning skills associated with *argumentation*. The types of thinking associated with argumentation are often incorporated into state standards and assessments. As a result, students must engage in inquiry procedures as they evaluate the quality of evidence and reasoning presented in support of a claim. Furthermore, standards from the National Research Council (1996, 2007) emphasize the need for students to make connections between explanations and evidence, and to take ownership and responsibility for their decisions.

Various authors have explored components of argumentation. For example, Toulmin, Rieke, and Janik (1984) defined *argumentation* as "the whole activity of making claims, challenging them, backing them up by producing reasons, criticizing those reasons, rebutting those criticisms, and so on" (p. 14). This is an important activity in the development of scientific literacy. Wallace, Hand, and Yang (2004) contended that an essential characteristic of scientific literacy is the ability to evaluate a scientific knowledge claim. To do this, students must understand the relationships between questions, data, claims, and evidence. This is the guiding mindset of our project, resulting in instructional procedures, an associated graphic organizer, and an embedded strategic approach to evaluation of claims and arguments.

This material is based upon work supported by the National Science Foundation under Grant Number 0554414. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

J.A. Bulgren (✉)
University of Kansas, Lawrence, KS, USA
e-mail: jbulgren@ku.edu

M.S. Khine (ed.), *Perspectives on Scientific Argumentation*,
DOI 10.1007/978-94-007-2470-9_8, © Springer Science+Business Media B.V. 2012

Purpose of the Study

By way of response to the challenge of helping students engage in higher-order thinking associated with argumentation, the purpose of this study was to develop a set of instructional tools that would support science teachers in helping middle and secondary students improve their scientific argumentation.

The project developed an Argumentation and Evaluation Intervention (AEI) and associated graphic organizer, the Argumentation and Evaluation Guide (AEG). We designed these materials to assist students in the following argumentation activities: (1) identifying a claim presented in a written science-based report or from an inquiry activity, and analyzing the claim for qualifiers; (2) identifying evidence, labeling the evidence by type, and judging the quality of the evidence; (3) identifying the reasoning that allowed the claimant to make the claim based on the evidence presented, labeling the reasoning by type, and judging the quality of the reasoning; and (4) drawing a conclusion about the quality of the claim and explaining the reasoning that supported the conclusion, including presenting rebuttals or counter-arguments.

The design study included two parts: a quantitative and a qualitative study. The results of the quantitative study demonstrated the efficacy of the AEI for use by teachers in inclusive middle and secondary school science classrooms that contain students of diverse abilities, including students with learning disabilities and those who have been identified as gifted (Ellis & Bulgren, 2009). Analyses indicated that significant differences were found for mean total test scores in favor of the students who were taught with the AEI over students taught with a traditional lecture-discussion format. In addition, for three of the four subscale scores, significant differences were found for students in the AEI condition over the comparison condition. These were the second, third, and fourth subscales described earlier.

The design study is built upon the research and recommendations of Bannan-Ritland (2003) and Kelly (2004). Central to the iterative design process are activities such as analyzing what needs to be developed, implementing iterative development until solutions are developed, and analyzing the effect of the solutions on teacher and student performance data. The design and development process begins with informed exploration to understand the situation (Bannan-Ritland, 2003). Then, the development process in an iterative approach includes: (a) identifying the design principles, (b) operationalizing the target cognitive processes, and (c) balancing the theoretical model with real-world requirements. This is a cyclical process with prototyping and testing continuing until ease of use and intended functionality is achieved. This process includes iterative user-centered design, interaction analysis, and usability and feasibility analysis. The process also involves evaluating the impact, which includes feasibility analysis, fidelity of implementation, analysis of the effect of the instructional process and materials on students' and teachers' skills and understanding, and efficacy studies.

Theoretical Framework

The conceptualization for evaluating claims and arguments is based on the theories of Toulmin (1958), who defined the central components of *argumentation* as reasoning from grounds or data (evidence) to knowledge claim (conclusion) with warrants (links from the evidence to the claim with principles and underlying assumptions), possibly additional backings, and qualifiers and rebuttals. In addition, others have contributed research and commentary on argumentation. For example, Driver, Newton, and Osborne (2000) noted the need to emphasize the correctness of judgments about arguments in addition to the structure of an argument. Lawson (2003) agreed with Driver et al. (2000), but noted the need for attention to real-world issues related to argumentation as students evaluate their own claims and conclusions.

The evaluation component of the strategy was based on components of scientific thinking that Kuhn (1991) calls the "skills" of argumentation. Evaluation components include, among others, an appreciation of the role of empirical evidence (Kuhn, Amsel & O'Loughlin, 1988) and judging the credibility of evidence in terms of (1) reliability (Schauble, 1996), (2) experimental control (Koslowki, Okagaki, Lorenz & Umbach, 1989; Kuhn Garcia-Mila, Zohar & Andersen, 1995; Schauble, 1996), and (3) objectivity (Klahr, Fay & Dunbar, 1993; Kuhn et al., 1995; Penner & Klahr, 1996; Schauble, 1996). We incorporated these evaluation components into the instructional procedures for use by science teachers in inclusive general education classes.

This project also incorporated the research-to-practice interface and supports as discussed by Klahr, Chen, and Toth (2001). These involve the use of direct instruction to prepare students for evaluation of evidence, support for transfer and generalization to other experiences, the provision of strategic skills that help students acquire domain knowledge, the ability to evaluate one's own use of these skills as well as those of others, and the goal of raising new issues for future research. These components fit well with explanations about how people learn and how they construct knowledge (Bransford, Brown & Cocking, 2000), and also include intervention strategies suggested by Carver (2001).

This project can also be built on the work of others relative to designing education *curriculum materials*. In terms of designing educative curriculum materials to promote teacher learning, this proposal incorporates heuristics that help teachers engage students in asking and answering scientific questions and making explanations based on evidence. This process is supported by procedural guides and professional development activities. These guides and activities may be applied to a variety of chapters, texts, and curricula, such as promoted by Davis and Krajcik (2005). However, while others (Linn, Clark & Slotta, 2003) provide already-prepared content and context for students to analyze for scientific arguments, our approach provides guides for students and teachers to use with a variety of student-generated and teacher-identified arguments and claims. These claims may be found in a range of scientific sources as well as real-world issues.

Products of the Design Study

We produced two major products from the design study, the AEI and the AEG with its embedded reasoning strategy. These products are instructional materials that include strategic-thinking approaches to support science teachers in improving the science argumentation knowledge and skills of middle school and secondary school students. The authors and teacher-researchers collaboratively developed these products through the design study process over a period of 3 years.

Description of the Argumentation and Evaluation Intervention

Content Enhancements. The AEI builds on a collection of instructional resources called Content Enhancement Routines developed by the Center for Research on Learning at the University of Kansas (KU-CRL) (Bulgren & Lenz, 1996). The Content Enhancement instructional interventions help students to process complex information using higher-level thinking skills. Previous studies indicated that students of diverse ability levels can learn content information using Content Enhancement procedures and that teachers can learn the instructional procedures easily. Researchers have found statistically significant results in favor of students who received instruction using Content Enhancements Routines when compared to students receiving traditional lecture-discussion instruction. These results were found for a number of routines including those designed to enhance concept acquisition (Bulgren, Schumaker & Deshler, 1988), learning by analogy (Bulgren, Deshler, Schumaker & Lenz, 2000), and manipulation of content information such as making comparisons (Bulgren, Lenz, Schumaker, Deshler & Marquis, 2002). This project was designed to move Content Enhancement research to a focus on higher-order thinking associated with reasoning about and evaluation of argumentation—a logical progression in the line of research.

The graphic organizer. The graphic organizer is the AEG (see Fig. 8.1 for an example AEG). The AEG contains a flexible cognitive reasoning strategy (the Argumentation and Evaluation Strategy) that guides students as they evaluate the components of arguments made in support of claims. Specifically, the strategy consists of the following steps: (1) identify the claim and qualifiers; (2) identify the evidence presented; (3) identify the type of evidence as data, fact, opinion, or theory; (4) evaluate the quality of the evidence; (5) explore the reasoning that connects the evidence to the claim; (6) identify the type of reasoning as theory, authority, or logic; types of logic include reasoning by analogy, cause–effect, correlation, or generalization; (7) evaluate the quality of the reasoning; (8) explore rebuttals, counterarguments, or new questions; and (9) draw a conclusion accepting, rejecting, or withholding judgment about the claim and explain the reasoning for the conclusion. We included supporting questions on the Guide as prompts for the learner. During the instructional process, students write information associated with each of the nine steps of the strategy on the AEG. The teacher guides the students to reach consensus on a class rendition of the AEG.

Argumentation & Evaluation Guide

Topic: Coffee and Health
Title: Coffee Drinkers Beware
Source: Research report from a funded project

Name: John A.
Class: Science
Date: 5-15-11

(1) What is the **Claim**, including any **Qualifiers**? Are there qualifiers? **Yes/No**. (If yes, underline them.)
Drinking coffee may cause heart attacks in sedentary people within two hours after drinking coffee.

(2) What **Evidence** is presented? In column 3, identify the type of evidence with the letter: **Data** (D), **Fact** (F), **Opinion** (O), **Theory** (T).	(3)	(5) What chain of **reasoning (warrant)** connects the evidence to the claim? In column 6, identify type of reasoning with the letter(s): for **AUTHORITY** (A), **THEORY** (T), or type of **LOGIC: Analogy** (AN), **Correlation** (C), **Cause-Effect** (CE), **Generalization** (G)	(6)
The University-based study of 500 subjects funded by a federal grant found that sedentary people were over 50% more likely to suffer a heart attack within 2 hours of drinking coffee than people in the general population who drank the same amount of coffee.	D	A cause-and-effect connection was found between sedentary people and heart attacks by a research study and a medical expert.	CE, A
The Principal Investigator, a Professor of Medicine, commented that this finding was likely to extend to the general population of sedentary people.	O	This means we can generalize the effects of drinking coffee to all sedentary people as a cause of heart attacks.	G

(4) **Evaluate** the quality of the evidence as poor, average or good. **Explain** your evaluation.		(7) **Evaluate** the quality of the chain of reasoning as poor, average or good. **Explain** your evaluation.	
Reliable	Good – large number of subjects	Strength of Authority	Good – respected sources
Valid	Good – used a controlled experiment	Application of Theory	Not present in article
Objective (no bias)	Good - Confirmed by independent doctor	Type of Logic	Good – cause & effect/generalization
Controlled Experiment - Yes			

(8) What are your concerns about the believability of the claim? (your counterarguments, rebuttals or new questions)?
I would like to see another big study. What is the risk for coffee-drinkers who are not sedentary?

(9) Accept, reject, or withhold judgment about the claim. **Explain** your judgment.
I accept the claim that drinking coffee may cause heart attacks in sedentary people because of good research data and the opinion of a respected medical authority, but I have more questions.

C Bulgren & Ellis, 2010

Fig. 8.1 A sample Argumentation and Evaluation Guide

The AEG graphic organizer provides a space for each step (and associated question) of the argumentation process, starting with step one (designated by the number 1) and the question, "What is the claim, including any qualifiers?" Then, step two prompts, "What evidence is presented?" There is an adjacent space for step three, "Identify the type of evidence" with letters that represent data, fact, opinion, or theory. Then, there is step four to evaluate the evidence, prompted by, "Evaluate the quality of the evidence as poor, average, or good." Next, there is step five to identify the chain of reasoning, with the prompt, "What chain of reasoning (warrant) connects the evidence to the claim?" Following step five is step six, a column for students to "identify the type of reasoning" with letters that represent authority, theory, or type of logic (i.e., analogy, correlation, cause–effect, or generalization). Then, there is a space for step seven to evaluate the reasoning, prompted by the challenge, "Evaluate the quality of the reasoning as poor, average, or good." Then, there is a step eight with the question, "What are your counterarguments, rebuttals, or new questions related to this claim?" Finally, in step nine, students are guided to arrive at their conclusion, "Accept, reject, or withhold judgment about the claim, and explain your judgment." In addition, a Scoring Rubric was developed for use in analyzing students' evaluation of a claim (see Fig. 8.2).

The instructional procedures. The AEI project materials support teachers by explaining instructional procedures (the Argumentation and Evaluation Routine

Argumentation and Evaluation Scoring Rubric

Name: ______________________ Date: ______________
Teacher: ____________________ Topic: ______________

Scoring Guidelines for each Step	0 Poor	1 Needs Improvement	2 Good Progress Toward Improvement	3 Very Good Meets Standards	Score
1 Claim	Student gives no response.	The student response inaccurately identifies the claim being made or writes a response not structured as a claim.	The student partially identifies the claim being made.	The student accurately identifies the claim being made.	
1 Qualifier	Students gives no response.	The student response fails to accurately identify qualifier(s) within the claim OR fails to state there are no qualifiers present.	The student partially identifies qualifier(s) within the claim that are present.	The student accurately identifies most of the qualifier(s) within the claim OR correctly states that none are present.	
2 Evidence	Student gives no response.	The student response identifies evidence that fails to support the claim.	The student accurately identifies some evidence used to support the claim.	The student accurately identifies most of the evidence used to support the claim.	
3 Identifying Types of Evidence: Data, Fact, Theory or Opinion	Student gives no response.	The student response fails to accurately identify any types of evidence.	The student accurately identifies some types of evidence.	The student accurately identifies all of the evidence as data, fact, theory or opinion.	
4 Evaluation of Quality of Evidence	Student gives no response.	The student response fails to accurately evaluate OR discuss the quality of the evidence.	The student evaluates and discusses some of the quality of evidence OR indicates that quality was not relevant.	The student evaluates and discusses the quality of evidence (i.e., validity, reliability, objectivity/bias or controlled experiment.)	
5 Chain of Reasoning (Warrant)	Student gives no response.	The student response fails to explain the author's reasoning connecting the evidence to the claim.	The student explains some of the author's reasoning connecting the evidence to the claim.	The student explains the author's reasoning connecting the evidence to the claim (i.e., authority, theory, or types of logic such as generalization analogy, correlation, or cause and effect).	
6 Identification of Types of Reasoning	Student gives no response.	The student response fails to accurately identify types of reasoning.	The student accurately identifies some types of reasoning.	The student accurately identifies types of reasoning (i.e., authority, theory, or types of logic such as generalization, analogy, correlation, or cause and effect).	
7 Evaluation of Quality of Reasoning	Student gives no response.	The student response fails to accurately evaluate the quality of reasoning OR explain his/her evaluation.	The student evaluates some of the quality of reasoning and/or explains some of his/her evaluation.	The student evaluates the quality of reasoning AND explains his/her evaluation (i.e., authority, theory, or types of logic, such as generalization, analogy, correlation, or cause and effect).	
8 Concerns of the student	Student gives no response.	The student response raises no new relevant concerns.	The student raises some new relevant concerns.	The student clearly raises new relevant concerns AND expresses them as counterarguments, rebuttals or new questions OR states there are none.	
9 Conclusion and explanation about the Claim	Student gives no response.	The student response neither makesa conclusion to accept, reject or withhold a decision about the claim NOR provides an explanation of his or her reasoning.	The student makes a conclusion to accept, reject or withhold a decision about the claim OR provides an explanation about his or her reasoning.	The student makes a conclusion to accept, reject or withhold a decision about the claim AND provides an explanation for his or her reasoning.	

Fig. 8.2 Rubric for scoring Argumentation and Evaluation Guides

[AER]) to use in instruction and discussion, and to guide student dialogue in whole group and small group cooperative structures. These methods have a sequence of three instructional phases: "Cue," "Do," and "Review." During the "Cue" phase, the teacher (a) introduces the scientific argument either through a reading or through reviewing the results of an inquiry investigation; (b) explicitly informs students about the importance and benefits of learning scientific argumentation; (c) distributes and explains a one-page graphic organizer, called the AEG; and (d) prompts the students to take notes on the guide and to participate in the discussion.

During the "Do" phase, the major part of the routine, all students complete all of the parts of the guide by following the set of nine strategic thinking steps; these are the steps of the strategy that are cued on the AEG described previously. The teacher scaffolds the process of learning to analyze and construct scientific argumentations by first developing the AEG as a whole-class, teacher-guided activity, second by having students work collaboratively on the AEG with substantial teacher guidance, sharing and discussing group work with the class to create a class version of the AEG, and finally to work in small groups to collaboratively construct the AEG and then to present and defend their AEGs to the class.

Finally, in the "Review" phase, the teacher reviews the information covered in the "Do" phase and the process that the students have used to analyze and evaluate the claim and supporting argument. In this phase, students reflect on their understanding of the process of scientific argumentation, and the teacher identifies areas for additional attention.

The teacher might begin the process by reading an article that includes a science claim and evidence (either the teacher or the students reading it aloud or students reading the article individually). The teacher can then engage in whole group instruction. In addition, the teacher has the option of forming students into small, collaborative research teams and guiding them through the process of constructing a graphic organizer to support the exploration, development, analysis, and evaluation of a claim.

The teacher always guides the students in the development of understanding by co-construction of the ideas within the class, beginning with a blank AEG and interactively constructing the AEG elements based on students' prior knowledge, insights, discourses, and explanations. It is important to note that the teacher *never* hands out this graphic study guide to the students in a completed form. Rather, the teacher completes a guide before the class only as an instructional plan to organize and clarify his or her own understanding, and the teacher and all students start with a blank guide and interactively discuss and complete the guide as a whole class or in collaborative groups. Therefore, the final guide might differ from what the teacher planned, because the teacher incorporates students' contributions, questions, and insights into the final construction of the guide developed through dialogue and a consensus-building approach.

The AEI is meant to support rather than replace the ways that teachers teach critical science content. Therefore, although the intervention includes components of direct instruction as recommended by Klahr et al. (2001) and Carver (2001), it must be stressed that the components of the intervention are flexible and do not replace hands-on inquiry experiences. However, the intervention does help teachers and students organize, synthesize, record, discuss, co-construct, and review understandings from a variety of sources and experiences, including the results of student science investigations.

Lessons Learned from the Design Study

We have separated the qualitative findings from this study into three general areas. First, the teacher-researchers provided insights into their views of argumentation, perceptions of their own abilities to teach argumentation, and their views about students' abilities to engage in argumentation. Second, we gathered information about classroom implementation of the AEI as teacher-researchers used the intervention in classroom contexts; this included suggested revisions to the instructional procedures. Third, we gathered information about general strategies that teachers used to support the instruction, and about the "big picture" insights into the cross-curricular use of the AEI.

Teachers' Views of Argumentation

Background. An impetus for this study came from reports that stress the need to help our students engage in higher-order thinking. Data from national assessments, such as the National Assessment of Educational Progress Report Card (NAEP) (Applebee, Langer, Mullis, Latham & Gentile, 1994), from the Program for International Assessment (PISA) (IES, 2007), and from some research projects (Kuhn, 1991; Means & Voss, 1996), have indicated that most young Americans do not have a firm grasp of higher-order reasoning such as that associated with argumentation. This is an important issue for all students in education today. However, according to the National Research Council (2007) in *Taking Science to School*, the norms of scientific argument, explanation, and evaluation of evidence differ from the norms students encounter in everyday life. As a result, an important goal in science education is that teachers are able to support students as they learn appropriate norms and language for productive participation in the discourses of science. This goal, in itself, is challenging. Added to this, however, is the need to teach a wide range of students of differing abilities in science classes. Students may include those who have disabilities, are gifted, or are average achieving.

This idea of science for all students fits with the views of the National Science Education Standards (NRC, 1996), of AAAS Project 2061 (1993), and a multitude of national reports calling for science literacy as a basic requirement for all citizens. The goal is that efforts to improve scientific literacy be infused throughout the K-12 curriculum for all students, not just for the best and brightest. Our design study addressed the challenge of developing instructional resources that promote the acquisition of higher-level thinking skills by all students. Unfortunately, all too often materials for academically challenged students focus only on lower-level knowledge and skills and deprive them of the opportunity to acquire the critical higher-level thinking skills required to engage in quality scientific reasoning.

Ysseldyke (2009) raised an issue of importance for this study related to quality science for all students. Ysseldyke contended that *all* students should be challenged. This would require a shift in focus from providing remediation for struggling students to making sure that all students struggle. He contended that all students should be working at what Vygotsky and Michael (1978) called their zone of proximal development. A critical feature of addressing the zone of proximal development is involving students in social learning with their peers through collaborative discourse on argumentation and evaluation.

Teachers' views at the beginning of the study. At the outset, after the authors had modeled and explained scientific argumentation to the teacher-researchers during a 2-week summer institute, the teacher-researchers agreed that higher-order reasoning associated with scientific argumentation is one of the most important scientific abilities. Furthermore, they agreed that it would help students respond to real-world needs, become informed citizens who are not gullible, and succeed on state and other assessments requiring higher-order analysis and evaluation. They, therefore, recognized that the skills of argumentation would easily impact multiple areas of science literacy. They recognized that students need enough content knowledge to engage in higher-order thinking, but also need general processes and procedures such as those used in the AEI.

Despite their general support for this type of instruction, the teacher-researchers were unfamiliar with details of the higher-order reasoning as it is characterized in Toulmin's approach to argumentation or with instructional strategies for supporting students in developing the knowledge and skills of scientific argumentation. They were open, however, to collaborating with the authors in a design study to investigate these issues.

Relative to teacher-researchers' views about their students, many of the teacher-researchers in the design study believed that some students were capable of understanding concepts such as those associated with argumentation, including reliability, validity, and objectivity. They were concerned, however, that many students were not mature enough to engage in this type of thinking.

Teachers' views at the conclusion of the study. By the end of the study, the teacher-researchers provided valuable information on all components of the AEI. Specifically related to claims and qualifiers, teacher-researchers, who taught classes ranging from sixth through ninth grades, thought that their students easily understood and found claims and qualifiers. One teacher-researcher indicated that, as the study progressed, students were becoming more aware of qualifiers to claims that they found in articles or infomercials outside of classes. However, teacher-researchers also indicated that some of the qualifiers that students found (particularly from outside sources) might well fit better as concerns, that is, rebuttals, counterarguments, or new questions.

By the end of the project, teacher-researchers reported that, relative to evidence (Toulmin's grounds), their students learned to distinguish between data, fact, theory, and opinion as types of evidence. An issue of concern for researchers, however, was that some teachers were still experiencing difficulty in providing clear explanations as to the difference in some evaluative components such as validity and reliability.

All in all, teachers believed that many students felt empowered in that they learned to think about a claim and were willing and able to develop questions about a claim or evidence and to organize their thinking. Thinking about the quality of evidence was particularly useful in that students thought more about reliability and bias. Students also raised issues about possible special interests or motivation of authority figures, even if those figures represented respected institutions. In this context, students raised the issue as to whether institutions might have vested interests in a claim due to grant support and funding.

Relative to the chain of reasoning (Toulmin's warrants), input from the teacher-researchers during their pilot of the AEI provided information on their beliefs about students' abilities related to higher-order reasoning that could link a claim to the evidence presented in an argument. During the development process, several teacher-researchers recommended that the intervention not use the more complex words on the AEG that were taken directly from theorists such as Toulmin. For example, they recommended using "chain of reasoning" to represent Toulmin's "warrants" and using "concerns" to represent "rebuttals." Although the researchers accepted these recommendations at the time, this ultimately raised issues regarding the wisdom of substituting some simpler synonyms for complex theoretical terms.

At the end of the study, when the teacher-researchers discussed the chain of reasoning, they believed that students seemed to understand how authority and theory served as appropriate warrants for a claim. However, the teacher-researchers believed that students had more difficulty with the complex area of logic. Relative to logic, they thought that students understood and used the term "logic" correctly in a general way, but did not understand various components of logic (as used in the AEG and in the instruction) such as analogy, correlation, causation, and generalization.

Relative to rebuttals and counterarguments, some teacher-researchers thought that these terms were difficult for students to understand, although they could more easily come up with new questions. One teacher-researcher thought that the greatest benefit came from student consideration of the last two components of the AEG: consideration of concerns and new questions, and drawing conclusions about the claims. Students, particularly in the upper grades, demonstrated some transfer of learning in that they commented on what they saw or read outside of class that contained claims. These included information found in infomercials, mailings, and various advertisements and articles.

One very important issue involves teacher-researchers' perceptions that students did not particularly enjoy the argumentation and evaluation instruction and activities. Researchers questioned whether this perception led some teachers to report that they would not use all the components of the intervention in the future. Student enjoyment is, indeed, a concern in education, but is only one consideration that must be subjected to more research.

Summary. In summary, this study challenged some of the teacher-researchers' prior beliefs about the level at which students can engage in higher-order thinking associated with argumentation. Some teacher-researchers made assumptions about the learning abilities of students in earlier grades, believing that they might not be able to acquire the higher-order thinking associated with analysis and evaluation of claims and argument. This was not borne out. In general, teacher-researchers valued the intervention and believed that it had a place in their curricula.

Implementation Issues Relative to the Argumentation and Evaluation Intervention

Background. This study is built on the work of others relative to designing curriculum materials that help teachers engage students in making explanations based on evidence, with a focus on the argumentation components put forward by Toulmin (1958). During the design study, however, teacher-researchers collaborated with the authors in attempting to modify the terminology and presentation of the Toulmin model in the AEG to adapt it to the prior knowledge, abilities, and experiences, especially of middle school students.

Implementation fidelity. Suggestions from the teacher-researchers regarding adaptations of terminology and instructional procedures raised issues related to fidelity of implementation. The strategy, graphic organizer, and instructional procedures are built on the research on Content Enhancements (Bulgren and Lenz, 1996). The work of the KU-CRL has emphasized the importance of fidelity of implementation. Rigorous research has been conducted on many of the Content Enhancement Routines, and much of this research has reinforced the importance of fidelity to the core components of research-based interventions.

The need for fidelity of implementation when using procedures that have been previously subjected to rigorous research is emphasized by Ysseldyke (2009) in his discussion of the importance of treatment integrity. He contended that when effective treatments are implemented with fidelity or integrity, the treatments have a strong effect on student outcomes, but that when that is not the case, research results can be misleading, making an observer think that a treatment is not effective when it was actually the implementation that was not effective.

Scheduling. Overall, the teacher-researchers recommended that the AEI be introduced early in the school year. They also recommended that the AEI be taught as part of the scientific process. This could help students recognize the need to look across science areas for competing claims in a variety of content materials.

Provision of examples. The teacher-researchers also recommended developing multiple types and examples of scientific and socio-scientific claims to include in a curriculum to facilitate integration of reasoning about claims and arguments across the school year. Examples suggested by the teacher-researchers included developing and/or analyzing scientific arguments that were based on published experimental studies, historical research, correlational studies (such as epidemiological studies in medicine), social-scientific claims, controversial issues (with multiple, competing claims), and student-generated results from classroom investigations.

Source of claims. Teacher-researchers used the intervention as envisioned, that is, with the written claims made by others, but also expanded its use to lab reports and provided other ways to improve the intervention. For example, teacher-researchers showed students how to write their lab reports based on the structure of the AEG as a way to clarify and present their own claim and arguments. Teacher-researchers found that the use of the AEG helped consolidate what students learned from labs. They also reported that just providing "hands-on" lab experiments did not ensure understanding by students of the critical science content related to the laboratory activity. By reviewing the AEG that students developed based on their experiments, students thought more about the experiment and how the results related to key science concepts.

Student discourse. The use of the intervention with students' own materials raised an important issue in terms of evaluation of their own results and collaborative discourse with other students. Even after students had conducted a lab experiment, they often found it hard to think of themselves as "authorities." Nonetheless, using the AEG allowed many students to challenge other students on their data analysis and accuracy leading to scientific discourse in the classroom. The teacher-researchers indicated that such a challenge might have seemed more objective (and perhaps less confrontational) when using the AEG as opposed to direct criticism. As a result, students seemed less likely to withhold comments on another student's thinking. One teacher-researcher thought that the AEG helped students most in lab reports because it provided supports for students to write summaries of the results of experiments and to write detailed justifications for their conclusions.

Terminology. Teacher-researchers suggested modifications to help students understand difficult or potentially difficult vocabulary. Therefore, an adaptation of terminology suggested by the teacher-researchers that we accepted during the study was the substitution of the word "concerns" for "rebuttals" or "counterarguments." Although well-intentioned, this suggestion by the teacher-researchers proved problematic. At the end of the study, students were interpreting the word "concerns" in a more personal way. That is, the issues they raised in that component of the analysis often included personal fears. Specifically, the problem that became obvious in the scoring sessions was that students often assumed that they should raise their own worries, rather than rebuttals, counterarguments, or new questions about a claim. This raises the issue that terms need to be clear and precise.

Sequential levels of implementation. All of the teacher-researchers provided incremental introductions to the AEG, sometimes breaking it into three parts and providing conceptual understanding of key vocabulary support at each level. These levels were as follows.

Level one argumentation emphasized the importance of initial learning activities focusing on the *big picture* of science argumentation. To do this, students were engaged in whole class and small-group discourse about their questions regarding the claim, about missing information in the argument, about rebuttals to the claim, and about students' concerns about the quality of the argument.

Level two argumentation engaged students in examining the *evidence* used to support a claim and in evaluating the quality of the evidence (which includes elements such as reliability, validity, and objectivity). Students gradually examined these elements of quality of evidence when questions arose in their discourse about the topics addressed in level one. Therefore, level two types of argumentation discourse often followed quickly and organically from level one.

Level three argumentation engaged students in examining the *chain of reasoning* (the warrant) that supports a claim. This examination includes identifying the type(s) of warrant used based on theory authority and/or logic. At this step, it is critically important that students can explain how the warrant connects the evidence to the claim, leading the reader to believe the claim. However, the teacher-researchers indicated that the use of the intervention required a great deal of time and experience for students to master these abilities.

Scaffolded materials. Other teacher-researcher suggestions related to the structure of the graphic organizer. This provided an interesting insight into the evolution of teacher thinking, perceptions, and analysis of instruction. For example, in the third year, the sixth-grade teacher-researcher indicated that using the whole AEG was overwhelming for her students. As a result, she broke the Guide into parts, put each of the parts on a separate page, and used each part, one at a time. Specifically, she prepared color-coded pages that contained the individual components of each section of the Guide. She reported that she initially believed that when younger students, were able to focus on the component parts individually, they would be able to better analyze the claim with its associated qualifiers, evidence, warrants, rebuttals, and conclusions. At the end of the study, when she probed student satisfaction with the color-coded guide, however, she found that the students, in fact, did not like the color-coded approach.

Another teacher-researcher projected the article onto a white board and let students choose colored markers to highlight parts of information that supported different sections of the AEG, which proved more appealing to students and seemed to encourage class discussion. These innovative additions, and students' responses to them, raised the issue of the need for teachers to engage in ongoing discourse with students to determine relatively quickly how students perceive the usefulness of innovative procedures. What may have face validity for the teachers may not have the same appeal to the students. These findings emphasized the importance of engaging classroom teachers as teacher-researchers in development projects.

Summary. In summary, although we encouraged the teacher-researchers in this design study to take a great deal of latitude in trying adaptations and modifications, their final feedback on how they would implement the intervention raises concerns as anticipated by Ysseldyke (2009) and others (Bulgren & Lenz, 1996). At the end of the study, some teachers indicated that they would continue to use only parts of the AEI. As indicated by Ysseldyke (2009), the "pick-n-choose" approach to the use of validated interventions does not lead to optimal student learning. This may become a very important issue for this intervention if teachers pick and choose which components to use and which not to use in the future, especially if they omit the higher-order thinking related to analyzing and evaluating evidence and reasoning. Alternatively, if teachers want to modify an evidence-based intervention, there must be ways to subject those modifications to the levels of rigorous research demanded by the field. An empirical question is the current level of fidelity of implementation on the part of teachers after professional development activities and the support that some teachers need to implement research-based interventions with fidelity.

Strategy Supports and Cross-curricular Use of the Argumentation and Evaluation Intervention

Background. A contribution of this project is the use of theory about domain-specific and domain-general knowledge and dimensions of scientific reasoning processes (Klahr et al., 2001). Scientific reasoning, according to Klahr et al. (2001), is classified by domain specificity versus domain generality as well as by the type of reasoning processes involved, such as generating hypotheses, designing experiments, or evaluating evidence; each of these three processes may be explored for either domain-general or domain specific knowledge. Furthermore, Reiser, Tabak, Sandoval, Smith, Steinmuller, and Leone (2001) contended that students must develop a deep understanding of science and use general strategies in particular scientific domains if they are to approach arguments more as experts than as novices. From another point of view, Stevens, Wineburg, Herrenkohl, and Bell (2005) argued for a move from fragmented approaches to more comparative and unified approaches in instruction that could, ideally, make school a meaningful place for students to learn and analyze even more complex, overlapping issues. This raises the issue of how widely applicable strategies that are considered general can be used to support a new intervention such as the AEI, as well as how applicable an intervention such as the AEI is in cross-curricular areas.

Strategies used in support of the Argumentation and Evaluation Intervention. Supporting strategies that teacher-researchers found helpful included those associated with questioning and reading. We found that teacher-researchers spontaneously used questioning in the classroom to scaffold the learning with the AEI. This highlights the importance of teaching supporting literacy strategies, such as those requiring paraphrasing and summarizing as needed. In addition, the use of questioning in instruction has been supported by meta-analyses. For example, Rosenshine, Meister, and Chapman (1996) focused on interventions that utilize questioning techniques in some form. They reviewed the studies focusing on teaching students to generate questions as a way of improving their comprehension during or after reading or listening to a passage. In general, they found that directly teaching students to ask and answer questions yielded significant differences in favor of the experimental groups with regard to tests constructed by the researchers.

Specifically related to our study, some question types that teacher-researchers found effective were those identified by Rosenshine et al. (1996). Among these effective questioning techniques, teacher-researchers often used signal words such as "why," generic questions, and question stems. To illustrate, the sixth-grade teacher-researcher reported that challenges occurred with the components of "chain of reasoning" (warrant). The Guide and instruction included three overall types of reasoning: authority, theory, and logic. The sixth-grade teacher-researcher provided scaffolding questions (with a "fill-in-the blank" format) for the students such as the following:

"Why does _____ *authoritatively* prove the claim?"
"Why does _____ *logically* prove the claim?"
"Why does _____ *theoretically* prove the claim?"

In addition, teacher-researchers used prompting questions that students could use as they explored rebuttals and counterarguments. These were questions such as the following:

"What *more* scientific information do you need?"
"Is there an argument *against* the claim?"
"What are questions for *further* investigation?"
"What more could be done to *improve* the research?"

Relative to reading supports for students who had reading difficulties, most teacher-researchers read the articles aloud to the class when articles were the focus of the activity. They also utilized "think-alouds" and modeling as they read the article to the class—particularly in the early learning phases of using the AEG. Various learning scaffold suggestions by teacher-researchers included modeling by the teacher-researchers how and where they themselves found the types of logic in the reading or in the laboratory activity.

Cross-curricular use of argumentation. Teacher-researchers provided insights into the cross-curricular use of a strategy focusing on the analysis of claims and arguments. At the outset of the study, the science teacher-researchers saw the goal of analyzing and evaluating claims and associated argumentation as specific to the domain of science in which they were teaching.

By the end of the year, feedback from the teacher-researchers and others in the schools indicated a broader value and use of the thinking skills associated with argumentation. Specifically, the teacher-researchers received positive support from others, including reading teachers, administrators, and other content area specialists. For example, both reading teachers and administrators who observed the use of the AEI indicated that this intervention was useful as a reading support as well as a science learning support at the eighth-grade level. The reading teachers and administrators pointed out the power of the intervention to focus on conceptual understanding of words such as reliability, validity, objectivity, theory, logic, and authority. They contended that deep understanding of these words would help students succeed on a variety of assessments.

Particularly important, the sixth-grade teacher-researcher provided support for the usefulness of the AEI as a good reading strategy. She reported that by the end of the year, students understood the need to read and re-read an article for different purposes—first for an overview, second to identify the claim, third to identify the evidence, and so on. This teacher also thought that it has helped students persevere as they read and re-read an article.

In general, the teacher-researchers believed that this intervention helped students read for deep meaning and big ideas rather than just searching for facts. As a result, students referred more to the text to understand components of an argument. Interestingly, teachers in other classes such as Advanced Placement Language Arts reported teaching procedures and processes similar to Toulmin's model in their persuasive writing curriculum. Therefore, they were very supportive of using this approach in science classes, and discussed coordination of terminology and goals across subject areas.

Summary. In summary, as the third year progressed, teacher-researchers shared insights about the potential cross-curricular power of the evaluation of claims and arguments. These insights arose in two areas. First, the teacher-researchers did not believe that they could effectively teach all that was needed about the evaluation of claims and arguments in 1 year, in one science class, and in one specific domain. Second, teacher-researchers shared new information about the objectives in their districts that had cross-curricular implications. Some noted that an emphasis on evaluation of claims and arguments was becoming important not only across science grades and courses, but that it also was becoming a focus in other areas such as Language Arts. This has led to suggestions for future research on the analysis of claims and arguments across domains and content areas. Research in this area would attempt to determine the components of the AEI that may be incorporated across subjects, domains, and disciplines.

Conclusions and Recommendations

As a result of this study, the AEI was shown to help students understand and use the components of argumentation as put forth by Toulmin. However, the study also resulted in recommendations for use of the intervention and future research.

Teachers need preservice courses and in-service support as they teach students to engage in complex higher-order thinking in science. For example, we observed some teacher-researchers struggling to explain the differences in concepts such as reliability and validity. Therefore, if national standards ask that students become citizens of the world by engaging in higher-order thinking, then the field must prepare teachers to have the background to help students do this. As a result, an empirical question for future research relates to the current preservice courses taught at the undergraduate level in universities, and the level to which they incorporate adequate support for teachers to engage in the teaching of higher-order thinking such as that required to analyze and evaluate claims and arguments.

Furthermore, it is possible that teachers would benefit from ongoing collaborative meetings with other teachers, both in their content area and in other content areas. In these meetings, ongoing collaborative discourse might well support innovative instruction associated with higher-order thinking within and across content areas.

Teachers also need ways to analyze and monitor their own impressions of how students are performing in their classes. Ongoing, effective and efficient ways to analyze students' readiness to learn, their perceptions of adaptations and modifications, and their enjoyment of the instruction are needed. For example, many teachers believed that younger students, such as the sixth graders in this study, were not able to learn to engage in higher-order thinking associated with argumentation. Students at that grade level, however, out-performed many other groups of students from higher grade levels (Ellis & Bulgren, 2009). In addition, a way to effectively and efficiently monitor students' perceptions of adaptations and modifications would be useful. This would have allowed the teacher who spent a great deal of time breaking apart and color-coding components of the AEG to adjust her modifications during, rather than at the end of, the project.

Another need is to explore ways to monitor students' views and enjoyment of the new instruction. Therefore, an empirical question relates to the correlation between student enjoyment and learning, and how students report their impressions. It might be that the use of student interviews rather than, or in addition to, objective satisfaction surveys in future research studies would provide valuable insights for teachers and researchers. In addition, future research is warranted on the use of formative assessments to determine not only student progress in using an intervention, but also students' views of the instructional procedures.

Teacher-researchers also raised issues concerning the fidelity of implementation of the intervention. Teacher-researchers, in general, indicated that they would not use the entire routine in the future, presumably because of the difficulties involved in higher-order thinking associated with argumentation—the very component that is being urged by researchers and commentators. This issue needs to be addressed by the field. When research findings indicate that a package of instructional interventions serves to help students learn, the value of fidelity of implementation must become an important issue in professional development and classroom use.

Therefore, future research is needed to explore the required levels of fidelity of implementation of the component parts of a research-based procedure to assure learning outcomes at levels similar to those found in the original research. Follow-up research into the effects of using only portions of components of a research-based intervention, rather than the complete set of components, would provide much-needed information for teachers and professional developers. If teachers want to modify a research-based intervention, there must be ways to subject those modifications to standards of rigorous research.

In addition, other research and development may be needed to incorporate enjoyable ways of learning, such as learning games, into instruction. The incorporation of such games may well add not only to student enjoyment of the learning process, but also to critical learning time and collaborative engagement for students.

A related need is to determine the number of times a teacher needs to implement a new intervention for students to benefit. For example, the ninth-grade teacher-researcher implemented the intervention only six times compared to 10 times by all other teacher-researchers, and her students performed at approximately the same level as the sixth-grade students.

This study also raised the possibility that interventions such as the AEI may have benefits beyond the specific area in which the original research was conducted. This is possible because of similar content literacy and higher-order thinking demands across content areas, subjects, domains, and disciplines. For example, it was reported that a reading teacher, after observing the implementation of the instruction in a science class, indicated that the AEI was one of the best exemplars of vocabulary development and support for conceptual understanding. She suggested that its use could have a positive impact on state assessments.

In addition, teachers from other areas, such as Language Arts, indicated that the higher-order thinking was the same as they emphasized in their courses when they taught persuasive writing. Components of the AEI that were considered useful across content areas included thinking about claims and the qualifiers to the claims, analyzing and evaluating evidence and reasoning, considering other options, and coming to and defending a conclusion about the worthiness of claims. Therefore, an empirical question is whether cross-curricular use of research-based instruction, such as in AEI, would enhance learning in different content areas due to multiple exposures to higher-order thinking challenges. Future research might also address the power of such interventions to improve student performance on state assessments.

In conclusion, the valuable contribution of both qualitative and quantitative studies on a single intervention contributed to a rich understanding of the complex challenges of teaching argumentation. Furthermore, the contribution of teacher-researchers in the classrooms provided valuable insights for the study as they used the AEI in regularly scheduled science instruction.

References

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

Applebee, A. N., Langer, J. A., Mullis, I. V., Latham, A. S., & Gentile, C. A. (1994). *The national assessment of education progress 1992 report card*. Princeton, NJ: Educational Testing Service.

Bannan-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher, 32*, 21–24.

Bransford, J., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain mind, experience and school*. Washington, DC: Committee on Learning Research and Educational Practice, National Research Council.

Bulgren, J. A., Deshler, D. D., Schumaker, J. B., & Lenz, B. K. (2000). The use and effectiveness of analogical instruction in diverse secondary content classrooms. *Journal of Educational Psychology, 92*(3), 426–441.

Bulgren, J. A., & Lenz, B. K. (1996). Strategic instruction in the content areas. In D. D. Deshler, E. S. Ellis, & B. K. Lenz (Eds.), *Teaching adolescents with learning disabilities: Strategies and methods* (2nd ed., pp. 409–473). Denver, CO: Love Publishing.

Bulgren, J. A., Lenz, B. K., Schumaker, J. B., Deshler, D. D., & Marquis, J. (2002). The use and effectiveness of a comparison routine in diverse secondary content classrooms. *Journal of Educational Psychology, 94*(2), 356–371.

Bulgren, J. A., Schumaker, J. B., & Deshler, D. D. (1988). Effectiveness of a concept teaching routine in enhancing the performance of LD students in secondary-level mainstream classes. *Learning Disability Quarterly, 11*(1), 3–17.

Carver, S. (2001). Cognition and instruction: Enriching the laboratory school experience of children, teachers, parents, and undergraduates. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 385–426). Mahwah, NJ: Lawrence Erlbaum Associates.

Davis, E. A., & Krajcik, J. S. (2005, April). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3–14.

Deshler, D. D., & Schumaker, J. B. (1986). Learning strategies: An instructional alternative for low-achieving adolescents. *Exceptional Children, 52*(6), 583–590.

Driver, R., Newton, P., & Osborne, D. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287–312.

Ellis, J., & Bulgren, J. (2009, March). *Professional development materials to develop student knowledge and skills of scientific argumentation.* (Year 3 report submitted to the National Science Foundation). Lawrence, KS: University of Kansas Center for Research on Learning.

Institute for Education Sciences—National Center for Educational Statistics. (2007). *Highlights from PISA 2006: Performance of US 15-year-old students in science and mathematics literacy in an international context.* NCES 2008016. Washington, DC: Author.

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *The Journal of the Learning Sciences, 13*(1), 115–128.

Klahr, D., Chen, Z., & Toth, E. (2001). Cognitive development and science education: Ships that pass in the night or beacons of mutual illumination? In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 75–119). Mahwah, NJ: Lawrence Erlbaum Associates.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*(1), 111–146.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: MIT Press.

Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation isn't enough: The role of causal mechanism, sampling method and sample size in causal reasoning. *Child Development, 60*, 1316–1327.

Kuhn, D. (1991). *The skills of argument.* Cambridge: Cambridge University Press.

Kuhn, D., Amsel, E., & O'Loughlin, M. (Eds.). (1988). *The development of scientific thinking skills.* San Diego, CA: Academic.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60* (4, Serial No. 245).

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

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education, 87*, 517–538.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2), 139–178.

National Center for Education Statistics (NCES). (2004). *Trends in international mathematics and science study.* Retrieved December 9, 2004, from http://nces.ed.gov/timms/Results.asp?Results=1

National Research Council (NRC). (1996). *National science education standards.* Washington, DC: National Academies Press.

National Research Council (NRC). (2007). *Taking science to school.* Washington, DC: National Academies Press.

Penner, D., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development, 67*, 2709–2727.

Reiser, B., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263–306). Mahwah, NJ: Lawrence Erlbaum Associates.

Rosenshine, G., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*(2), 181–221.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*, 102–119.

Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005, Summer). Comparative understanding of school subjects: Past, present, and future. *Review of Educational Research, 75*(2), 125–157.

Toulmin, S. (1958). *The uses of argument.* Cambridge: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning*. New York: MacMillan.

Vygotsky, L. S., & Michael, C. (1978) *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wallace, C. S., Hand, B., & Yang, E. (2004). The science writing heuristic: Using writing as a tool for learning in the laboratory. In E. W. Saul (Ed.), *Crossing borders in literacy and science instruction* (pp. 355–368). Newark, DE: International Reading Association.

Ysseldyke, J. (2009). When politics trumps science: Generalizations from a career of research on assessment, decision making, and public policy. *Communique, 38*(4), 6–8.