**Scientific Literacy in the Context of Civic Reasoning:**
**An Educational Design Problem**

Sharon Derry, University of Wisconsin-Madison
Daniel Zalles, SRI International
Annual Meeting Presidential Address
American Educational Research Association
April 2011

## I. INTRODUCTION

21st century cyber infrastructures for distributed learning, communication, and knowledge sharing are unprecedented in the opportunities they provide for global citizens to come together to address serious problems of our time. Science (both physical and social) and scientific research play critical roles in this discourse, as do the design features of the communication environments themselves. Yet collectively solving problems within communities will require going beyond science and design to help citizens develop a full range of practices and understandings that underlie responsible civic reasoning. Examples include leadership, collaborative competence, and argumentation. These, like scientific conceptual systems and communication technologies, are important cultural *tools* (Vygotsky, 1987) that mediate complex interactions within socio-technical environments. An important goal for science instruction and schooling in general is to help learners become responsible fluent users of these complexly interwoven cultural tools (Wertsch & Kazak, 2011). Without such fluency, the collaborative work made possible by technology cannot reach its full potential.

New technologies supporting instant mass communication and organization can empower positive forms of collaboration and activism. Yet, all too often, these new technologies simply aggravate disrespect, confusion, and bias. For example, technology-enabled public discourses on issues of importance to our world and communities, which are absent the cues and norms of face-to-face communication that may reinforce courtesy and respect, frequently become banal and addictive venues in which participants vent frustrations but fail to engage in productive argumentation or consensus building. We illustrate with excerpts from hundreds of public online discussions of recent events in Madison WI, edited to capture the flow of the discourse as succinctly as possible. The excerpts are from an online discussion of a Wisconsin State Journal article reporting a University of Wisconsin professor's scientific economic analysis of Governor Walker's Budget Repair Bill.

<u>Badger backer</u>: *I have seen a number of studies to say this will help. Why should I trust yours?*

<u>Wackwack</u>: *No, badgerbacker12, you haven't seen even ONE study which suggests Walker's plan will help. If you had, you would've shared a link. Link, or you're a liar.*

Rural Resident: *An opinion based by some model used by a UW Madison professor, isn't that part of the reason we've fallen subject to bubbles and bubbles bursting???? . . .*

Gohawks girl: *Of course the bill will set Wisconsin back . . . My household will pay the increase in pension and healthcare contributions, approx $400 a month. We readily acknowledge that increased contributions were needed, but it still takes $400 a month out of our pocket. We will trim what we spend on everything. Multiply $400 a month by 65,000 state workers in Dane County and you get more than $260,000,000. That's all kinds of businesses not getting revenue, that's sales tax not being collected, rent and mortgages that are late, small businesses that have to close their doors . . . If Wisconsin becomes a right to work state, consider the effect of businesses cutting wages for hundreds, maybe thousands of employees. New jobs at $8 an hour can't pay the rent . . .*

Stilllaughing: *Scooters plans ARE creating jobs people. Look at the growth!! Attorneys, credit counselors, repo men. I thought he would just be creating McJobs, . . . wrong . . .*

Peewee: *Obama just launched over 100 missiles into Libya and the left-wing ignores it and continues to whine about entitlements. lol*

Lext: *Pewee -- Who says we're not upset about . . . Libya? As I recall, this article wasn't about that issue. I don't know what would prompt someone to bring that up here.*

Harvey: *. . . I believe their true agenda is to keep reducing revenue coming into the government so they can, as Grover Norquist said, 'starve the beast' and shrink government so it can be easily disposed. They want us to become a true capitalist society which is a good thing IF you believe corporations operate for public good.*

Gorman: *Badgerbacker 12---it's been almost three hours now. What are the other "studies" you have seen?*

Somebody: *Private industry will feel this. As a public employee I will make less each month, this will hurt the economy. I will also boycott employers who supported this take away of rights, benefits and pay. I will spend nothing of what I have left on businesses that supported taking that away from me and my family.*

Jabee: *Keep up the boycotts. Let the rich buy Johnsonville brats and Koch tp and Menard's stuff. There are plenty of better products out there so don't buy from the WI traitors. They are already blanching at the loss of business and hoping that voters forget, but they are wrong. Even those of us who have not lost income and are not union workers refuse to support Walker and his cohorts . . . Let the recalls roll.*

Green Man Rising: *If you do not think the boycott is having an effect you are wrong. Kwik Trip is offering the first lbs of bananas for free. Milk is really cheap. They want you to forget they backed Scotty. I take their free bananas and low priced milk to make sure they keep losing money. I fill up across the street. Purchase only what is on sale from the backers, and shop the rest somewhere else. Johnsonville brats are not jumping off the shelf anymore. Keep up the pressure.*

Fink 1025: *I am sure that Johnsonville Brats were not jumping off the shelves three weeks ago either. It is not exactly grilling weather. I DO NOT support Walker or his bill, but I will not boycott the businesses. The employees are the ones hurt by these boycotts the most.*

Cents: *Bogus argument. Agree government employees will have less income, but if we don't reduce costs, we need to increase taxes. If we increase taxes, ALL of us have less disposable income . . . Thus, there will be less disposable income either way; however, at the level of government employees pay and benefits, they will still have plenty to go out to eat, movies, vacations, buy clothes, etc. Some of the people paying taxes to support government employees (i.e. middle class) do not have that luxury and just hope to save their house, or be able to buy gas to get to work, etc. Thus, Walker is doing the right thing. . .*

Muir: *. . . the best article I've seen all day in my news trawl today has been this cap time interview . . . http://host.madison.com/ct/news/local/health_med_fit/article_2888e8be-4fee-11e0-81f8-001cc4c03286.html Bobby Peterson said: They fear a population that is educated . . . Under-educated and unknowledgeable is exactly the kind of staff Walker is comfortable with - they play at his level . . .*

Ace4: *Dear Muir, The only thing larger than your inflated opinion of yourself is possibly the utter lack of fiscal restraint. What a pompus #### you are.*

Our analysis of the longer discourse of which these excerpts are part did *not* indicate that critical thinking was absent. On the contrary, several lengthy arguments with some supporting evidence were submitted. However, these arguments were often imbued with emotion and personal bias, and they were almost always disparaged or, more frequently, seemingly ignored by other participants in favor of shorter, pithier statements. And very few people owned what they said by using their real names.

*IA. Scientific Civic Reasoning as an Educational Goal*

We propose that *active, collective citizenship through responsible civic reasoning, empowered by tools of science and technology, is an important educational goal of our time.* To that end, we advance a proposal, a feasible strategy that, if widely adopted, might help improve civic reasoning with and about science. The approach we suggest is based on a tradition that values knowing how to *use* science, rather than just knowing about science, and it resonates strongly with the argument that the purpose of school is to prepare students for future learning (Bransford & Schwartz, 2002). There is precedent in prior research for expanding the scope of classroom science instruction to encompass a broader pantheon of authentic scientific activities. For example, Ford and Foreman (1986) identify three broad parameters of science as a social, material, and rhetorical practice. However, our vision looks beyond disciplinary boundaries of the typical science classroom toward a re-conceptualization that encourages spread of a future cultural practice: responsible scientific civic reasoning.

We acknowledge that generating productive instructional discourse in science classrooms and then making connections between that and future public discourse is an extremely complex matter. In practical terms there are many hurdles to overcome, including deeply ingrained historical, physical and administrative structures that hamper interdisciplinary thinking. They include deep cuts in educational funding in states like Wisconsin, Michigan and California, which strain resources, hinder innovation, and increase the likelihood of underprepared and overworked teachers in the nation's classrooms. And they include NCLB-spawned accountability policies that favor individual rote and basic skills learning over critical, collaborative and creative thinking.

Roadblocks to success also include disturbing recent statistics such as a finding reported by Michigan State University (2007) that 70 percent of all U.S. adults struggle to read and understand the science section of the New York Times. This general lack of scientific literacy may explain the dismal state of scientific discourse found on public blogs and discussion boards, but it also illuminates the complexity of our task in theoretical, as well as practical, terms. The process of mastering the cultural tools of scientific civic reasoning is fundamentally social. Vygotsky (1987) made this point by arguing that higher mental functioning appears first on the "intermental" then on the "intramental" plane:

> When encountering a new cultural tool such as a statistical instrument, this means that the first stages of acquaintance typically involve social interaction and negotiation, between experts and novices as well as among groups of novices. It is precisely by means of participating in this social interaction that interpretations are first proposed and worked out and hence available to be taken over by individuals. (Wertsch & Kazak, 2011)

As Wertsch and Kazak explain, one property of cultural tools is that they are robust in allowing for understanding at many different levels while supporting discourse that can promote development. Participants use words to communicate even when they have very limited understanding of the complex ideas underlying them. It is this paradox that allows them to enter into discourse with more informed others -- including teachers, scientists, and community experts -- thereby leveraging their way toward increasing levels of understanding. This view illuminates the importance of citizen scientists engaging in problem solving with other citizens. For example, Hacking (1975) described the historical role of scientists and mathematicians in mentoring the lay public regarding probabilistic reasoning. But citizens who are not professional scientists, acting in their roles as stakeholders and decision makers within affected communities, must be prepared to listen and learn. And professional scientists must also listen and learn in order to understand and appreciate the local and cultural contexts in which scientific reasoning takes place. In fact, all participants must become citizen scientists, albeit with varying degrees and types of scientific and local expertise. Helping achieve this capacity for deliberative discourse among citizen scientists is our vision and goal for science education at all academic levels.

In light of this complexity, a modest proposal may seem naïve. And yet we are optimistic that our strategy, if widely disseminated, might help individual teachers, schools, districts, or university programs think about how their current instruction can be re-designed, sometimes with fairly modest adjustments to current practices, in ways that help boost expectations for and quality of public discourse among future citizen scientists. Our approach begins by raising awareness of an interdisciplinary model for conceptualizing a pantheon of skills and understandings, which we refer to broadly as civic reasoning. We will overview and illustrate the characteristics of civic reasoning, discussing their implications for design of science instruction in the context of today's unprecedented global challenges and technological opportunities. To help build stronger connections between science and civic-mindedness, we propose that educators reflect on, evaluate, and revise their curricula and practices to require and develop the full range of skills involved in civic reasoning. We also suggest that educators evaluate and share practices representing how discipline-based science can be taught in ways that interact with larger societal phenomena. Toward that end, we provide cases from our own work and that of others to illustrate the design of science instruction based on a full model of civic reasoning.

### *IB. Six Phases of Civic Reasoning*

The dimensions of civic reasoning and societal issues around which a new set of interdisciplinary learning goals could be directed need further elaboration through discussion and research. For now, we offer a beginning framework consisting of six interacting phases of civic reasoning that attempt to characterize how, ideally speaking, societies can take action to identify and solve problems that have science components. We acknowledge that this idealization is more or less adhered to, depending on the circumstances. Although participation in each phase both requires and stimulates scientific literacy, we hypothesize that the phases differ from one another by emphasizing different skill sets and forms of reasoning. The six phases are briefly described as follows:

1. *Seek understanding, common ground, and consensus around what constitutes problems that are worthy of research and funding*. Some key capabilities required for this phase include: ability to analyze controversies, to understand and empathize with multiple conflicting views, seeing their similarities and differences; metacognitive awareness of bias related to self interest and other sources; ability to frame problems as researchable questions; and collaborative discourse skills, including the ability to negotiate tradeoffs and compromises.
2. *Leverage power structures to assure funding and support for important avenues of scientific research*. Some key prerequisites for success at this stage include motivation to support research for a cause; scientific literacy related to that cause; knowledge and resourcefulness related to identifying funding sources; and leadership and communications skills that include the ability to frame and evaluate arguments, persuade and influence.
3. *Design and carry out scientific research*. Prerequisites for this phase include deep scientific knowledge related to research problems; ability to formulate hypotheses and operationalize variables and meaures; technical analytic capabilities; and the skills and habits of systematic, disciplined inquiry.
4. *Practice peer critique in scientific communities of practice.* Important prerequisites for work in this phase include expertise in the scientific discipline of inquiry and dedication to a culture of principled, unbiased, constructive critical discourse.

5. *Use evidence to set policy or take other forms of civic action*. This phase focuses on influencing and developing action strategies that respond to or implement scientific findings. It requires ability to collectively tie problem analysis to options for action that represent tradeoffs (such as cost versus benefit) in a world with multiple competing priorities. Skills of analysis, negotiation, decision-making and leadership are among those required.
6. *Evaluate policy effectiveness.* This phase requires technical knowledge related to designing, conducting, and analyzing evaluation options, and communication skills needed to disseminate findings.

The six phases draw and build from the notion that students' sense of belongingness to their human and natural environment (Freire, 1970/1995) motivates them to become more engaged in the pursuit of technical knowledge and skills (Grueneweld, 2003). Three roles characterize citizens as stakeholders in their "place": (1) recipients of changes that impact the place, for better or worse; (2) technical contributors to the well-being of the place (e.g., professional scientists and other careerists); and (3) leaders and decision-makers about ways to improve the place.

Of these three positions, only the first is an inevitable role that all citizens must play by virtue of their residence in their place. The second role changes its characteristics depending on the needs of the place because it is dependent on the shifting nature of what technical knowledge is necessary among the workforce to further the betterment of society. The third role is taken up by individuals who emerge from the citizenry to take elite leadership positions, yet for the betterment of democratic citizenship, it can certainly be expanded to encompass a broader view of the ordinary citizen as a contributor to civic betterment through more informed position taking and more productive and beneficial civic action. Grueneweld (2003) takes this position when he advocates for ecological place-based education as a way for the citizenry to contribute to current and future environmental improvement.

It is this concept of what constitutes active participation beyond merely contributing professional technical assistance in the area of expertise that constitutes the basis for our argument that classroom science learning outcomes would be both greater and more contributory to social betterment if framed within a larger pantheon of civic decision-making and action to which all may eventually contribute effectively and of which all will be impacted, for better or for worse.

There is already precedent for a broadening of the concept of what it means for ordinary people to engage in place-based science through the "citizen science" movement, where ordinary citizens contribute place-specific data to professional science researchers. For example, the website for the National Oceanic and Atmospheric Administration elicits citizen involvement in research, stating: “Across the United States and its coastal waters, opportunities exist for volunteers to take part in research, observation and educational roles that benefit science, our citizens and our planet (http://www.volunteer.noaa.gov/index.html).

Yet, these acts of citizenship are largely focused on contributing to the pool of technical knowledge and data rather than pursuing science as understood within a larger framework of civic reasoning within which scientific research plays a critical yet not exclusive role.

Hence, we believe that attention to all six stages is likely to improve interest in science because it makes more explicit the connections between the science and the individual learner's personal life. In addition, we believe that it is likely to improve understanding of science and scientific inquiry by showing how science interacts with larger societal pursuits. An added benefit will come through how it contributes to the greater social good by modeling and asking learners to practice more socially productive and personally fulfilling civic deliberation grounded in scientific reasoning.

### *IC. Using a Civic Reasoning Model in Instructional Design and Evaluation*

We recommend that educators use a full model of civic reasoning as a guide for thinking about, discussing, redesigning, and evaluating science instruction at all grades and levels. The model can be usefully applied to understanding and examining what is emphasized or what is missing within a single lesson or course, within an entire curriculum, or within the program of study pursued by any individual student.

For example, in pursuit of the first phase of civic reasoning, instructional programs can be devised that get students to (1) recognize and articulate the controversies surrounding the issues that drive scientific research, (2) understand arguments that frame the controversies, (3) struggle with whether a particular issue being addressed by scientific research is truly, in the student’s view, a problem, or a greater problem than other problems, (4) exercise self-awareness of the origins of their own positions in factors such as personal or collective-self-interest, religious or ethical principles (5) build greater empathy for sources of others' views, and (6) channel that empathetic understanding into deeper appreciation for what societal stakes frame the scientific research about the issue.

We build  on typical constructivist assumptions regarding science instruction, “that students’ background knowledge profoundly affects how they interpret subject matter and that students learn best when they apply their knowledge to solve authentic problems, engage in “sense-making” dialogue with peers, and strive for deep understanding of core ideas rather than recall of a laundry list of facts” (e.g., Windschitl, 1999, p. 751). However, we contrast our approach with typical constructivist practices in which students are encouraged to ponder solution options critically yet rationally, under the assumption that the focal situation is accepted consensually as constituting a “problem.” Although constructivist approaches are a step in the right direction, they often do not go far enough. The real world is messy not only because problems do not have unequivocal solutions. It is also messy because, with some unequivocal exceptions such as natural disasters, one person or one group’s attribution of what issues constitute problems may not be another’s. The identifying of what constitutes a problem in the real world is often not amenable to social consensus and is only partly the result of rational reasoning. Values, emotions, social pressure, and perceptions of individual or collective self-interest compete with rational reasoning in the pantheon of contributors to decision-making about what constitutes a problem in the subjective mind of the beholder. This reality underscores the need to attend to other phases of civic reasoning.

Our recommendations are predicated on studies of decision-making in loosely coupled organizations (Cohen March, and Olsen, 1972; March and Olsen, 1979), and it is fair to characterize a society as such. We as members of society try to ponder solution options on a particular scientific issue rationally, yet we often ignore the critical framing component pertaining to how much the "problem" is perceived as such by all. As a result, decisions to enact strategies that are offered as solutions have only relative amounts of connectedness to differing perceptions about what constitute problems. All issues are likely to be controversial, though some more so than others, and ultimately, power play, conflict, and negotiation (Bardach, 1977) move agendas forward. For any particular issue however, the power-oriented components that drive agendas forward have the potential to be diminished for the better when more productive models of group deliberation are modeled and practiced, thereby exposing the lines of reasoning and valuation that characterize how the participants in the deliberation struggle with the issue, and laying the groundwork for the building of consensus.

Our model and recommendations are just as applicable to higher education, at both graduate and undergraduate levels, as they are to K-12 education. We will first illustrate this idea broadly, followed by more explicit examples from our own work in K-12 and higher education, where we have attempted to expand the foci of our programs into addressing fuller sets of civic reasoning phases than are typically addressed.

*K-12 Science Education.* An examination of national and state standards indicates that most skill-building in K-12 science falls in phases 3 and 4, which are about accumulating canonical scientific knowledge and designing and carrying out discipline-appropriate scientific research, with some attention also to persuasively communicating and critiquing scientific results. Other phases, if covered at all, are typically addressed in standards for disciplines taught in separate courses, such as social studies. Such disciplinary isolation provides a myopic look into the world of science wherein scientists are merely sets of discipline specialists rather than citizens immersed with everyone else in the broad interdisciplinary struggles that challenge us.

*Graduate Education.* One of us is a professor of The Learning Sciences at the University of Wisconsin-Madison (UW-Madison). The Learning Sciences field purports to be interdisciplinary in precisely the ways represented by these phases. Yet, if we surveyed such programs, would we find an overwhelming predominance of attention to phases 3 and 4, just as in K-12 science? It would be instructive to know the extent to which individual graduate students in different learning science programs engage in work that integrates across phases and the extent to which the programs encourage this integration. There is precedent for such an expanded interdisciplinary approach in post-secondary education. For example, Stanford University's Science, Technology, and Society (STE) program is built around the idea that key educational objectives are "Understanding the natures, causes, and social consequences of scientific and technological developments, how science and technology function in different societies, and how social forces attempt to shape and control these forces to serve diverse, often conflicting interests." (http://sts.stanford.edu/qa.html).

In sum, we argue that a transformative interdisciplinary approach to science teaching that addresses the full model of civic reasoning will be required in order to: (1) increase the citizen's understanding of interrelated scientific, moral, economic, and cultural aspects of world problems; (2) enable citizens to practice reflection and engage in productive discourse about what problems exist and what actions to take; and (3) incorporate into public discourse the methodological and interpretive challenges of evaluating effects of policies and actions. We propose that educators use a civic reasoning model to guide their thinking and discourse about science instruction at all grades and levels. For example, an exercise for an academic departmental retreat might be to examine the department's curriculum in terms of how much it contributes to the problem of civic reasoning. The model can be usefully applied by one teacher to understand and examine what is emphasized or what is missing within a single lesson or course, or within the program of study pursued by individual students, such as the graduate students advised by a particular faculty member. We argue that the model is just as applicable to higher education, at both graduate and undergraduate levels, as it is to K-12 education. What follows are more explicit examples, presented as three cases from our own work.

## II. CASE 1: HAL ONLINE

### *IIA. Overview*

Human Abilities and Learning is a large upper-level foundations learning sciences service course offered by the Department of Educational Psychology at the University of Wisconsin-Madison and required by many majors, including teacher education. The course addresses the scientific basis of thinking and learning and what this implies for guiding children and adults, for personal development, and for building environments that help people learn and grow successfully. It is designed for teachers and future teachers broadly defined to include parents, professionals, volunteers and others whose lives involve mentoring, guiding, teaching. Typically the course is offered as a large lecture course containing 150-180 students. However Derry offers a non-traditional section of this course (HAL Online) for students preferring an innovative problem-based learning format that meets face-to-face several times during the semester but is taught predominately online. There are 45 students enrolled in spring semester, 2011. HAL Online makes use of the Moodle course management system and aims to develop scientific literacy through reading and intensive discourse around problem-solving situations in which students apply ideas from the learning sciences. Students are assigned to small collaborative groups of 3-5 members that have common interests and majors and that work together, mostly online, throughout the semester.

The units in the spring 2011 HAL Online offering are:

I. Cognition and Culture, which focuses on topics such as the cognitive basis of learning, remembering, and the relationship between thought and language; research regarding cultural influences on thought; and models for “good thinking” such as hypothesis testing and argumentation logic (Halpern, 2003; Kuhn, 2005; Toulmin, 1958).
II. The Amazing Learning Brain, which focuses on early brain development in language and mathematics, expertise and lifelong learning, and the brain science of mindfulness (Blakemore & Frith, 2005; Jensen, 2008; Coles, 2004)
III. Using Learning Science in Reflective Practice, focusing on the promises and problems of constructivist teaching and on developing teachers’ skills at designing constructivist learning environments and analyzing learners’ thinking and problem solving within those environments (e.g., Eagan, 2010; Hirsch, 1996; Windschitl, 1999).

Each unit comprises four or five week-long lessons. During a lesson, students read and access multimedia resources about science content. Readings and other resources emphasize the cognitive foundations of critical thinking and neuroscience for educators and are drawn from textbooks and from video and news sources such as TED.com and The New York Times. In alternate weeks students either post a reflective personal blog that answers a problem-solving prompt, or participate in online collaborative problem-solving tasks requiring use of science content. Students are required to demonstrate literate use of the science content in their blog posts and discussions, which are graded using a rubric that rewards understanding and intelligent use of target ideas. There is an individual essay quiz after two units and at the end of the course.

The course has been developed over several design iterations and there is ample evidence for its success. It is typically well rated by students. Two studies using validated instruments and blind scoring compared learning in HAL Online with a set of matched control students drawn from lecture sections of the course. In both studies students in HAL Online very strongly out-performed matched control students in their ability to apply ideas emphasized in all course sections to reason about thinking observed in videos of children’s problem solving (e.g., Eagan, 2010). These results replicate findings of a similar study comparing traditional and online problem-based learning (PBL) sections of educational psychology courses taught at Rutgers University (Derry et al., 2006).

### *IIA. Application of the Civic Reasoning Model to HAL Online*

The following are two example lessons that will help illustrate how the phases of civic reasoning are used to analyze and think about lessons from HAL Online. These two lessons occur early in the course and supply students with models and norms for scientific thinking that will hopefully scaffold their learning and discourse throughout the course. As the following analysis will attempt to illustrate, the forms of reasoning promoted by the two different lessons apply to different phases in the civic reasoning model.

***Lesson A: Evaluating a Speaker’s Argument to Determine Whether to Invite a Proposal for Funding:***

Lesson goals: To acquire psychological concepts and skills that will enable you to:

1. analyze and judge the quality of arguments made by others, including the arguments of politicians, pundits, friends, relatives, and classmates.
2. improve your own thinking and ability to persuade by understanding and being aware of the quality of the arguments that YOU are making to yourself and to others as the basis for decisions and beliefs.
3. help improve the learning in your small-group online discussions (and beyond) by making sound arguments to support your ideas and by holding others accountable for the quality of their arguments.

Activities and Schedule:

I. Before noon, Feb 11 (Fri):
1. Study Analyzing Arguments -- Halpern Ch 5
2. Take online self-check quiz.

II. From noon Friday, Feb 11 - Monday, Feb 14: Participate in small-group online discussion task. Use the following TED video on educating African leaders: http://www.ted.com/talks/patrick_awuah_on_educating_leaders.html

Online Discussion Task: Should We Invite a Proposal for This?

Your group is the advisory board for the Gates Foundation, which has pledged substantial aid for Africa. Your group is screening presentations by leaders who want to make proposals for funding. Many preliminary presentations are being made and only 5-10% of them will ultimately be funded.

Patrick Awuah has presented his work that focuses on developing higher education for African leaders (see forum video). He would propose to expand the number of Ashesi U campuses in Ghana to reach more students. His proposal will include a study to evaluate the impact of his project.

Is the presentation convincing enough to warrant an invitation for full proposal? Full proposals require substantial time and resources to prepare and evaluate, so only promising ideas should be invited. Make a recommendation based an evaluation of Awuah's argument.

One person should volunteer to start the work with a draft analysis of Awuah's argument. All members should respond to this analysis with suggestions and questions that will help improve the group's evaluation. Before Monday night another person should volunteer to summarize the group's recommendation and reasons, addressing the question: Should The Foundation request a full proposal from Patrick Awuah? There is no set length for your final analysis and recommendation, but think less than a page.

***Lesson B: Designing an Evaluation Study***

Lesson Goals:

To deepen your understanding of a universal scientific cultural model for good thinking. You should understand (be able to explain and identify examples of) the following ideas, including their interrelationships, and be able to use them as tools for reasoning and for critiquing the reasoning of others:

1. dependent and independent variables
2. deductive and inductive reasoning
3. sample size, generalization, and related bias in reasoning
4. correlational and experimental evidence and their relative usefulness for making causal claims.
5. operationalizing and measuring variables

Activities and Schedule:
I. Before noon, Fri Feb 11 (Fri):
1. Study Thinking as Hypothesis Testing, Halpern, Ch 6
2. Take online self-check quiz.
II. From noon Friday, Feb 11 - Monday, Feb 14: Participate in small-group online discussion task. Use the following TED video on educating African leaders: http://www.ted.com/talks/patrick_awuah_on_educating_leaders.html

Online Discussion Task: Writing an Evaluation Proposal

UNESCO has agreed to fund a study by a team of social scientists (your group) to evaluate Patrick Awuah's causal hypothesis about the effects of Ashesi University (see video on educating African leaders). Your team wants this job and must prepare a proposal on how it will conduct the study.

Using ideas from Thinking as Hypothesis Testing, especially the section on three-stage experimental designs, your team should design and propose a realistic (feasible) study that would shed some light on the value of the program even if it were not able to answer the big question of whether leaders that think critically would improve a whole continent.

One person should volunteer to start the work with a draft proposal. All of you should respond to the proposal with suggestions and questions that will help improve it. Before Monday night, another person should volunteer to summarize the group's proposal. There is no set length for proposals but think less than a page.

*Applying the Model to Understand and Compare Lessons A vs. B.* Lesson A addresses phase 1, reaching consensus on what constitutes an important problem for research, phase 2, learning about processes involved in funding research, and to some extent phase 5, taking action based on available evidence. Related to phase 1, the problem asks students to decide whether a presentation makes a convincing argument that a particular problem related to educating African leaders is worthy of possible research funding by a major philanthropic foundation. Related to phase 2, students experience and learn about a legitimate process that simulates activities of a preliminary proposal grant review panel. Related to phase 3 students are asked to weigh the evidence given by a speaker and take action by either inviting or not inviting a full proposal.

In contrast, the lesson of example B requires design of an evaluation study to determine the success of an educational program for African leaders. Students must consider and debate such issues as formulating research hypotheses, operationalizing variables and measures, and experimental design. This task is thus related primarily to phases 3 (conducting research), 4 (critiquing research within a community of practice), and 6 (evaluating policy).

*IIB. Analysis of Student Thinking Processes*

*Method*. Derry examined how thinking processes of students differed for the two HAL Online tasks, confirming that they did focus on targeted phases of civic reasoning. In spring semester of 2011, 45 students registered for HAL Online were assigned to small groups of 3 or 4 members with similar majors or interests, and these small groups were randomly assigned to either section A or section B of the course. Each section comprised six small groups. For the first 3 weeks students in the both sections received the same assignments, readings, and were guided online by the same two instructors. In week 4 the students in section A received the lesson of example A that emphasized phases 1 (problem evaluation); 2 (funding research) and 5 (taking action) of the civic reasoning model. Groups in section B received the lesson of example B emphasizing phases 3 (scientific research), 4 (critique of research) and 6 (evaluation design). Several aspects of lesson design that were unrelated to the focus on phases were held constant for experimental control. For example, both tasks employed an identical TED video as stimulus material and the online guidance provided by instructors was loosely scripted (for example, students in both sections were encouraged by instructors to justify their decisions). Derry examined the online discourse for 6 of the 12 groups, three in each section. The following summarizes some findings from this analysis.

*Thinking in Lesson A: Evaluating a Speaker's Arguments to Determine Whether to Invite a Proposal for Funding.* All groups in this section conducted a responsible, thoughtful, reasoned discourse. The task was challenging, but students seemed enthusiastic about it. Good leadership emerged in all groups. All groups engaged in viewing and reviewing the video to understand the speaker's claims and evidence supporting them. Two groups formally evaluated the speaker's argument in terms of how well its premises supported the conclusion the panel would have to reach, to invite a full proposal. The other group evaluated how well the speaker supported his main hypothesis:

> *This is hard!! I started to analyze this argument and will continue to break it down to understand. Here is what I have so far:*
> *1. There is at least one premise and one conclusion. One conclusion is that: educating leaders in Africa is fundamental*

> *to it's ability to transform. A premise for this is: with an education focused on ethics and critical analysis of problems, leaders will be able to make better decision for the economy and people of Africa. . .*

As a prelude to making their decisions, all groups explicitly identified and discussed assumptions, argument fallacies, and implied arguments in the speaker's presentation. All groups grappled with the question of whether the argument was convincing by debating whether the evidence brought by the speaker was strong or weak and whether it was logically linked to the conclusions requiring support. Of particular concern was the value and validity of the speaker's narrative evidence provided as stories from Africa based on the speaker's personal experience:

> *You mentioned the storytelling, Liz -- I too wasn't convinced that the anecdotes were the most convincing support for his conclusion. During the beginning of the speech when he described the power outages at the hospital, I couldn't tell what point he was trying to make. When he tied it back into leadership by saying that some basic better management would have averted these problems, it then made sense. However, I question whether that anecdote constituted support for the conclusion. Those two incidents with power outages might exemplify the lack of planning and quality control in Ghana's health care -- or they might have been isolated, atypical occurrences. Is quality of patient care really suffering due to poor leadership? What concrete measures of health care are available? Stronger support for the conclusion might be found in the form of numerical data reflecting the shortcomings of hospitals.*

No two argument analyses will agree exactly. The examination of the students' work indicated that some groups did not identify all possible argument components and logical connections that the researchers saw. All of the group's analyses could have been improved through additional feedback and mentoring. But although they were not perfect, the students' decisions were reasoned, thoughtful, defensible and detailed, showing good to very good understanding of the tools of argument and the speaker's argument. Overall, the effort and intellectual depth of the students' work was impressive. Two groups supported inviting a proposal and one group conditionally supported a proposal if the proposer could sufficiently answer several questions.

*Thinking in Lesson B: Designing an Evaluation Study.* The groups in this section also engaged in a responsible, reasoned discourse in which every member contributed to and responded to others. The content of the discussions in these forums dealt largely with practicalities of conducting a large-scale evaluation. For example, all groups considered and then two groups dismissed the possibility of conducting a large-scale study with random assignment to schools. Due to real-world moral and budget constraints two groups opted for a non-random (nested) correlational design. All groups extensively discussed ways of operationalizing and measuring dependent variables. All groups grappled with issues of control and bias, recognizing the usefulness of procedures such as blind scoring, random assignment, and deceiving subjects about the purpose of the study. Only two of the groups made explicit reference to the talk in the video: One group examined the video to determine what the speaker was trying to argue, to help formulate the scientific hypothesis for their study. Another group studied the speaker's conceptualization of critical thinking and ethical leadership as a basis for designing outcome measures.

> *I think the proposal is a great start to the prompt. But I feel that in order to accurately measure for leadership skills, decision making, and work ethics these phrases might need to be defined according to operational definition. Since the proposal mentions that the participants will be measured on issues in Ghana and the world, is that part of the definition then as to what defines having leadership? I do not exactly have a definition for these terms but just curious if it would help the experiment if they were defined.*
>
> *While re-watching the video, I understood more of what was going on here and I realize that it would be beneficial to go off of the definitions for leadership and work ethic that Patrick Awuah gave. He states that leadership is not "political leaders" but rather "the elite, those who've been trained, those whose job is to be guardians of society." As for work ethic, he defines it as having "passion for what they're doing, the persistence, their ability to deal with ambiguity, their ability to tackle problems unseen before." I think having heard this, the evaluation and assessment in the proposal makes sense and will ensure that what we are testing is found. It is good to have an interview to see how people answer problems unheard of before.*

No group participating in lesson B challenged the value of the speaker's ideas on education or the quality of his argument. They did not find the argument problematic and spent relatively little time studying it. They focused entirely on the technical and practical aspects of how to design a study that might evaluate the speaker's educational claims scientifically. They considered the strategies, details and practical and financial constraints involved in collecting scientific, statistical evidence. They likely developed a deeper appreciation for how difficult and expensive it can be to obtain scientific evidence. Their ideas about what is possible sometimes bordered on naïve, although in some cases naïve ideas were successfully challenged by an instructor or other students and evaluation strategies were adjusted to better fit realities.

In contrast to the groups receiving lesson B, lesson A groups examined the video extensively and tried to understand the speaker's argument before judging it worthy or unworthy of further consideration. An interesting outcome of discussions emerging from lesson A was that students became interested in the shortcomings, strengths and meaning of personal stories as evidence. They struggled with drawing inferences from stories to conclusions. Importantly, they spontaneously began to reflect on the differences between scientific and narrative evidence, which relates to Kuhn's (1998) distinction between genuine and pseudo-evidence. This is an important distinction for students to understand because personal stories are widely used in public arguments and can be powerful in persuasion. They are used more often than statistical evidence and may often be the only type of evidence available. Personal experiences have validity as evidence if their limitations are recognized and they are appropriately questioned and interpreted. Students in group A appeared to advance their understanding of evidence by reflecting on this issue.

In sum, there seemed to be meaningful differences in the types of reasoning required and promoted by the two lessons, and the lessons likely led to different learning outcomes. Those designing a study to evaluate an idea did not deeply examine, *a priori*, the quality of the problem and related hypotheses but did come to understand much about the realities of designing studies to obtain scientific evidence from real-world program evaluations. Those grappling with the value of an idea for further consideration as a funding proposal more deeply examined the worthiness of the problem and potential payoffs of further funding and research. They also deeply considered the nature of evidence and the difference between narrative and scientific evidence. Although the students valued scientific evidence, the lesson A problem did not engage them in exploring any technical, ethical or practical issues regarding how to obtain it for the research they might fund.

## III. CASE 2: PRACTICING DEMOCRACY THROUGH DELIBERATIVE DISCOURSE

### *IIIA. Overview*

In pursuit of figuring out ways in which teachers can instigate in their students the first phase of civic reasoning (i.e., the seeking of common ground and consensus around what constitutes problems and needs) Zalles created Practicing Democracy through Deliberative Discourse (P3D), an instructional process for teaching and learning about controversial issues. P3D is grounded in the assumption that constructive, generative discourse around controversial issues requires skills that can be taught, modeled, and practiced. The issues may be strong or weak on the following dimensions: (1) high implications in terms of self-interest, (2) high moral implications, (3) high stakes for the community (as defined by geography or other parameters), and (4) high conduciveness to rational deliberation about costs and benefits in relation to what is known and not known.

The classification of how strong or weak any particular issue is on these dimensions requires a subjective judgment for which there is not a right or wrong response. For example, some may judge the issue of whether genetically engineered foods should be grown as primarily a matter of personal health (are they just as nutritious as natural foods?), economic wellness (is the practice cost-effective?), or morality (is humankind exceeding its proper place in nature to produce such foods?). This act of classification is the first step in the P3D process because it establishes the meta-cognitive and reflective character of issue deliberation. In the case of genetically modified foods, somebody framing the issue as primarily one about nutritional value or cost-effectiveness would likely want to engage in rational deliberation about evidence. Yet, somebody framing the issue as primarily moral may be closed to rational arguments about costs and benefits and either completely shut out rational arguments or seek to find a morally acceptable middle ground between total support and total prohibition of the practice.

The other key opening activity in the P3D instructional sequence is that of individual students' building knowledge about the issue by responding with an argument to a simple question that asks if they agree or disagree with a particular proposition. The classification exercise may either precede or follow this phase, depending on the amount of prior knowledge they have. If the students have enough prior knowledge about the issue to do the rating, the rating process can be the first activity in the sequence. The task of selecting the question is quite important because it influences the characteristics of the arguments that students will generate in response. For example, when deciding how to catalyze argumentation on the topic of the minimum wage, the question "Should the government mandate a minimum wage" intersects with argumentation that is likely to arise from "Do minimum wage laws hurt small businesses?" or "Are minimum wage laws necessary to ensure fairness in the workplace?" Yet, though they intersect, the questions are not parallel.

In the next stage of P3D, students read pro and con arguments about the issues and respond to a set of questions designed to get them to take preliminary positions and reflect on why they have taken their positions, which they then carry with them into group deliberative discourse. Below are the questions they are asked to respond to and in italics, for the sake of scaffolding, are examples of possible responses to the question of whether to ban violent videogames.[1]

1. Now that you have read the pro and con arguments, are you for or against (the debated course of action)?
2. Did any of the points made in the arguments help you decide? Which one(s)? Which points were most persuasive?
3. What outcomes do you expect if the (the debated course of action) is implemented? *For example, you may expect that violent crime is likely to go down if violent videogames are banned.*
4. What outcomes do you expect if the (the debated course of action) is not implemented? *(For example, you may expect that violent crime is likely to stay the same or go up if violent videogames are not banned.)*
5. Do your ethical values or beliefs influence your position? How? *For example, your values may dictate to you that violent video games should not be banned because that would be a violation of free speech.*
6. Does your sense of self-interest influence your position? How? *For example, your self-interest may dictate to you that violent video games should not banned because you would not be able to play them anymore and you enjoy playing them.*

[1]These examples of possible responses were added to the P3D materials after the piloting described in the next section, where the banning of violent videogames was one of four issues students addressed.

7. Does your connection to groups you identify with (e.g., social, cultural, ethnic, religious geographical) influence your position? How? *For example, your interest in the future of society may lead you to conclude that violent video games should banned because the antisocial or criminal behavior that can arise from widespread playing of these games could threaten public safety)*
8. What other knowledge or experiences of yours influences your position? (*For example, your knowledge may dictate to you that violent video games should be banned because you believe from evidence that violent video games may lead to antisocial behavior among people already predisposed to it.)*

One intention of P3D is that, through these exercises, students become sensitized to how subjective is the very idea that particular issues are problems, and to the relativity of each issue in terms of how high-stakes it is.

Unlike in a traditional debate, where the goals are to argue competitively and single-mindedly for a point of view, then pick a winner, the position taking that the individual learner is asked to do in P3D is not an end in itself but only a conversation starter for the seeking of common ground in the deliberative discourse that comes next in the instructional sequence. Students come together in small groups to share how they responded on their assigned issue, and then make preparations to report out as a group on the following:

- What did you agree about (i.e. what was your common ground?)
- What did you agree to disagree about?
- What follow-up actions does the group want to take (such as conducting a survey or other research study, visiting somewhere relevant, reading more about the issue, being an advocate for your position).

### *IIIB. Pilot Test of P3D in a High School Setting*

*Zalles* pilot-tested P3D among a class of 11th-graders in a small private school located in a small city (population 44,265) in a large ethnically diverse rural agricultural area of California. Demographic data are unavailable but Zalles was informed verbally that 10 to 20% of the students have learning disabilities. Many of the students are relatively affluent however and college preparatory.

The P3D process is adaptable in the sense that it can be applied to many different issues or sets of issues. In the piloting 11th grade classroom, where the primary goal was to get students to build their communication and self reflection skills, students were given many choices for what issues to deliberate about. The original intention was for the classes to be split into four groups and for each group to focus on a different issue that was distinguished from the other issues for being especially highly rated on one of the dimensions.

The pilot took place in January and April 2009. At this pilot, the framing of driving questions was constrained by what concise published pro and con arguments already existed on the issue on a student debate web site known as Debatepedia. On the topics capital punishment, violent video games, immigration, and human-induced carbon dioxide emissions, Zalles and the teacher of the students decided that the driving questions would be (1) Should capital punishment be prohibited? (2) Should violent videogames be prohibited? (3) Should greater numbers of immigrants be allowed into the U.S.? and (4) Should the U.S. government do more to regulate amounts of carbon dioxide emitted by people into the atmosphere?

### *IIIC. Analysis of Student Thinking Processes in the High School P3D Pilot*

*Method*. Based on detailed field notes, Zalles examined how the students in the high school pilot test responded to the four issues. Students worked in small groups and reported to their classmates about their deliberations. There were many interesting outcomes, some of which could be foreseen and some not.

*Capital punishment group deliberation*. The group that deliberated about whether capital punishment should be prohibited found common ground in the fact that each of their arguments were primarily about the morality of capital punishment, as opposed to other possible arguments such as a sociological one that capital punishment may deter crime. Where they diverged was on what moral argument they used. One student said he supported capital punishment but the other three students said no. The student who said yes said that by permitting capital punishment, people are playing God and that is immoral. The other three felt that some crimes are so heinous that the perpetrators deserve to be executed. The primary consequence of their discussion was common endorsement of a new policy whereby murderers are given a choice of life imprisonment or death. This would allow for the death penalty to be instituted and circumvent the morality problem of people playing God by taking another person’s life against their will. Various objections were raised to this argument by their classmates, including (1) that by giving criminals this choice you are giving them more freedom than they deserve; and (2) that this policy would be a form of euthanasia, and any form of euthanasia is morally wrong.

*Violent video games group deliberation.* In contrast, the group that deliberated about whether violent video games should be banned framed their positions in rational terms about society's interest rather than in the moral or personal self-interest terms that they could just as well have adopted. They exchanged viewpoints about whether banning violent video games would be a deterrent to social problems that may arise as a result of people playing them. The outcome of their discussion was strong interest in gathering more information about whether there are causal links between violent video game playing and antisocial behavior. Another outcome of their discussion was mutual agreement that whatever information they gathered would need to be weighed for persuasiveness against dueling priorities of trying to protect free speech while at the same time also trying to protect public safety. One student raised an unforeseen socio-economic argument that banning violent video games would be economically disastrous for the people who depend on the violent video game industry for their livelihoods and that the welfare of these people should be taken into account. What was notably absent from these students' deliberations was any attention to the notion that a ban might be wrong because the students themselves enjoy playing violent video games. What was also notably absent among the students was any argument to the effect that exploiting violence for the purpose of entertainment may be immoral.

*Immigration group deliberation.* The group which deliberated about whether the U.S. government should allow more immigrants to legally enter the United States found common ground in mutual endorsement of the notion that this is a very complicated and confusing issue which requires that they learn more about it to become more definitive. The arguments raised by this group contained a combination of economic and moral justifications, and attention to morality predominated. One student for example argued that immigrants should be allowed to settle in the United States because all people have the moral right to seek the best lives for themselves. This argument was countered by a pragmatic yet non-U.S. centric argument that too much immigration into the U.S. is bad for the countries from which the immigrants are emigrating because some of these people may be their home countries' best and brightest, hence their emigration threatens their countries' abilities to make the sort of improvements that are necessary to stop even more people leaving.

*Human induced carbon dioxide emissions group deliberation.* The group which discussed whether the U.S., government should do more to regulate human-induced carbon dioxide emissions was the only group within which there was complete agreement on the driving question. All group members agreed that more should be done to slow down carbon dioxide emissions and all came to the conclusion that the arguments they read that questioned whether carbon dioxide emissions are causing global warming were too "biased" to take seriously.[2] This was an interesting outcome because they did not in contrast perceive any bias in the counter arguments they read. Their disagreements were primarily around tactical challenges of formulating policies that would provide the best solutions to the problem. For example, would carbon dioxide emissions be decreased more through voluntarism or through regulations? Also, how much should be regulated and how quickly? They reported that the most consequential outcome of their discussion came after they read one of the published arguments in which a controversial claim is made that human-induced carbon dioxide is at most only a minor cause of global warming when compared to contributions rendered by solar activity and naturally-emitted carbon dioxide. However rather than persuading them that hence there should not be much regulation of human-induced carbon dioxide, they concluded that all environmental problems need to be tackled, even if humans are not the biggest contributors, because small changes made by many culminate in big changes.

### *IIID. Discussion of P3D High School Pilot Results*

[2] It is noted that the students were not asked to explain why they felt the capital arguments were biased. This would have been an appropriate line of questioning in hindsight.

Through the P3D process, these applications of classroom deliberative discourse have the potential to catalyze better learning outcomes by training students to become more engaged in science as citizens. This is done by embedding the learning and appreciation of science within the process of grappling with controversial issues. It is useful to conjecture how each of the four 11th grade student groups would respond if they were asked to reach consensus on the extent to which science would play a role in their further engagement on the issue. It is likely that the capital punishment group would not see a big role for science because their framing was primarily around morality. The videogames group might gather empirical data looking for evidence of correlations between playing violent video games and violent behavior, yet not render a judgment about videogame policy without factoring in moral and social concerns about personal freedom and public safety. In contrast, the immigration group would be less likely to engage with science because their arguments were morality-based rather than focused on economic or environmental impacts of increased immigration. Yet, this group's confession of not feeling knowledgeable enough to feel definitive about their position, either pro or con, and their strong interest in building up their knowledge about the issue would be a strong indicator of the possibility that their research would reveal new knowledge to them about economic or environmental ramifications and that this new knowledge would engage them in the physical or social science surrounding the issue. The topics of discourse among the group that focused on carbon dioxide emissions would readily lend themselves to further scientific engagement in two directions; first, in the direction of socioeconomic studies about the impacts of voluntarism versus regulation on behavior; and second in the direction of studies looking at cumulative impacts of carbon footprint-reducing actions that individual citizens can take.

These results among other things bear witness to students' abilities to admit that they need to learn more about a topic before becoming too definitive and that there is value in conducting research studies as a vehicle for this learning. Yet, these results also provide evidence of how people can quickly dismiss as "biased" arguments that they were not already inclined to agree with.

*IIIE. Pilot test of P3D in Teacher Professional Development Setting*

P3D also received a different form of piloting as a professional development workshop which was attended by 31 teachers from four diverse school districts serving 29,274 students, of which 25% are English-language limited and 9% have Individualized Education Plans (National Center for Education Statistics, 2011). There were equal numbers of teachers teaching fifth, eighth, and eleventh grade American history. These teachers were provided the opportunity to subsequently use the materials and training provided at the workshop to implement P3D with their students during the 2011-12 school year. The P3D workshop was one of a series of monthly workshops on different American history topics that the districts are offering to these teachers under a U.S. Department of Education Teaching American History Grant.

The goal of the workshop was set by the school district: to train teachers to do a better job teaching their students about the history of American immigration policy up to the present day. They conjectured that this goal would be more achieveable if the teachers adopted the P3D process as the way to teach about this topic because students would get to see themselves as the latest in a long line of citizens who have struggled with this issue throughout American history and hence, with this personal connection, build up greater empathy for past Americans and greater interest in the topic. Zalles customized P3D materials to focus the teachers on getting their students to practice deliberative discourse about whether legal immigration[3] is good for the United States. The district coordinators and Zalles agreed that deliberative discourse on this topic would best be followed by immersing students in arguments that have been used by distinguished Americans past and present (such as John F. Kennedy) to support or oppose immigration.

Zalles began the workshop by overviewing the P3D process. In his overview presentation to the teachers, he introduced a new exercise designed to sensitize the teachers to the subjectivity inherent to rendering a judgment about the extent to which any facet of society constitutes a problem, and then, how much of a problem compared to others. Zalles asked the teachers to respond to the question, "Socially speaking, which of these are problems, solutions, both, or neither?"

- Terrorism
- Cancer
- Legal immigration
- Illegal immigration
- Endangered species
- Genetically modified foods
- Taxes
- Capital punishment
- Social networking

Then they were asked to revisit the question but from a personal rather than social perspective. Finally, they were asked to rate the same issues for how high stakes each issue is.

Zalles then asked the teachers to split into groups by grade level and practice the entire P3D activity sequence, starting with reading pro and con polemics that appear in respectable published books about whether legal immigration is good for the United States. The polemics Zalles chose represented a cross-section of cultural, social and economic arguments concerning the immigration issue. Some were written recently and others were written many years ago yet still resonate today. Subsequently, the teachers practiced the remaining stages of P3D (described in the Case 2 overview section ) by filling out individual responses to the arguments, meeting in groups to practice deliberative discourse, and reporting out to the other groups. This was then followed by each of the three grade level groups observing on videotape how the immigration group in the prior high school pilot (just described in the previous section of this paper) responded to the same issue. The purpose of this exercise was to get the teachers to see

[3] The focus on "legal" immigration rather than "illegal immigration" was deliberate because Zalles and the district coordinators wanted the teachers to consider the goodness of immigration rather than the law enforcement issues surrounding what to do with illegal immigrants. The high school pilot test also focused on legal immigration.

that students could successfully engage in the process on the same topic that they just engaged in. After this, the teachers met in their grade level groups to plan how they would adapt the process to the specific needs and abilities of the students in their grades. Lastly, they reported out to the entire group. Each teacher was encouraged to implement a suitable adaptation of the P3D immigration unit in their classes at some point during the remainder of the 2010-11 school year.

*Method*. Prior to the workshop, the teachers were given a preassessment that was designed to provide baseline data about their understanding of P3D as an instructional method. At the conclusion of this school year's set of workshops, the districts will administer the same questions to the teachers to gauge development of understanding. The figure below presents the assessment items and scoring criteria:

**Practicing Democracy through Deliberative Discourse**
**Immigration Issue**
**Teacher Professional Development**
**Pre-Post Assessment**

**1. Pose three questions that have driven pro and con arguments about immigration past and present. You may start each question with "Is immigration ..." or "Does immigration...?"**
**(Example: if this item were about the issue of campaign finance reform, an appropriate question might be "Does unregulated campaign financing contribute to political corruption?" or "Is the regulation of campaign financing a violation of the constitutional right of free speech?").**
*Assessment Construct*: Understanding the underlying themes of the U.S immigration debate, past and present
*Possible good answers* would be whether immigration depresses wages, takes away jobs from other Americans, threatens the language or culture of America, introduces greater crime or threatens national security

**2. Rank each of your three questions from #1 on how challenging it would be for people with opposing viewpoints to find common ground in how they respond to the question. Then explain your rankings.**
*Assessment Construct:* Understanding the characteristics of typical sources of position taking among ordinary citizens on controversial issues and applying that understanding to the immigration controversy.
*Good answers* will (for example) sufficiently identify how the themes grounded in economics arguments are more amenable to rational discussion, research, and potential common ground building than are cultural arguments that may boil down to whether cultural diversity is better or worse for the U.S.

**3. Compare and contrast how you would score a student debate on immigration compared to a student deliberative discourse on immigration.**
*Assessment Construct:* Understanding the fundamental difference between the way success is defined in a debate compared to the way it is defined in a deliberative

discourse.
*Good answers* will identify that success in a debate is usually defined by how well the debater presents his or her argument and who on that basis "wins" the debate. In contrast, success in deliberative discourse is defined by how deep and reflective are the responses of the individual discussants to the controversial issue, how well they search for common ground through the exercise of good listening and questioning skills, how well they can communicate what they agree and disagree about, and what follow-up actions would be appropriate to advance the discourse.

## IV. CASE 3: DATA SETS AND INQUIRY IN GEOSCIENCE EDUCATION: ASSESSMENT BY THE MODEL

### *IVA. Overview*

In pursuit of figuring out ways in which high school teachers in particular can catalyze among their students phases 3, 5, and 6 of civic reasoning, (designing and carrying out scientific research, using findings to set policy or take other forms of civic action, evaluation of policy effectiveness) Zalles and other SRI researchers carrying out the Data Sets and Inquiry in Geoscience Education (DIGS) project (NSF GEO 0507828) designed and pilot tested an online-administered unit and performance assessment about investigating local climate change. The goals of the DIGS project were to produce innovative ways to structure supplementary curricula and assessments around common high school geoscience topics using appropriate technologies and real data sets. In addition to this unit and assessment about climate change, DIGS project collaborators at the Concord Consortium produced a unit and assessment about crustal plate boundaries and earthquakes using real data from the U.S. geological service. All DIGS materials are available for classroom use at http://digs.sri.com.

The objectives of the 5-6 class period climate change unit and 1-2 class period performance assessment are to have students examine real publicly available data about whether the climates of two particular cities, Phoenix and Chicago, are getting warmer and why. In the process, students are asked to formulate evidence-based arguments from the data and then recommend policies and policy evaluation strategies. The Phoenix climate is the focus of the instructional unit and the Chicago climate the focus of the performance assessment.

Students use Excel data tables and geographic visualizations to critically investigate authentic temporally and spatially-distributed data sets about climate change in one city (Phoenix) in relation to environmental problems. Then in a 2-day near-transfer performance assessment, they investigate a second city, (Chicago) with data from the same public sources, and compare and contrast the two cities. The tasks require that the students analyze and synthesize unfiltered data about climate change in order to determine if and why the climates of the two cities are warming (Quellmalz and Zalles, 2010; Zalles, Quellmalz, Gobert, & Pallant, 2007). In the unit, students draw conclusions about whether multiple decades of temperature data about Phoenix suggest that a shift in climate is taking place or whether the data show nothing more than typical short-term natural variability. Students also compare the changing trends in Phoenix to larger geographically-distributed temperature trends. Then, they investigate if there is evidence of a relationship between the temperature data and data that would suggest human influences, think critically about what can and cannot be known from the available data and propose an ideal follow-up research study for society to carry out. These tasks prompt students to apply inquiry strategies as they respond to data-related challenges such as (1) taking a position after reviewing conflicting claims about the same data, (2) recognizing the difficulty of rating something without sufficient comparative context, and (3) deciding how much evidence is needed to be convinced of something.

The performance assessment task walks students through a set of parallel procedures and questions which required that they transfer skills and understandings they accumulated in the unit to a near transfer task. The assessment requires that the students apply the methods and findings from the investigation of climate data in Phoenix to Chicago, a different city with dissimilar characteristics. The assessment was designed to be carried out by students as individuals whereas the unit was designed to encourage group interaction under the assumption that group interaction would be more conducive to greater learning. The assessment also distinguishes itself from the instructional unit by posing questions in formats that are more likely to yield valid interpretable data about student attainment of targeted skills and understandings than would the formats of the questions in the instructional unit, which were designed to induce divergent thinking and multiple perspectives.

### *IVB. Pilot Test of the DIGS Unit and Assessment*

The DIGS unit and assessment on investigating local climate change was pilot tested with 100 diverse students from two high schools in the San Francisco Bay Area. The school site of the first pilot test served a largely homogeneous community. The school had a high percentage of White, non-Hispanic students (75%) and Asian/Pacific Islander students (15%). Only 2% of the school's students were Hispanic. The setting of the pilot test at the school was an elective environmental science course for 11th and 12th graders. The teacher incorporated the Climate module into a unit she was already teaching on global warming.

In contrast to the setting of the first pilot test, the school setting of the second pilot test served an economically diverse set of communities. What differentiated the second pilot school from the first was that the second had a much higher percentage of Hispanic students (38%) and smaller percentages of White non-Hispanic (38%) and Asian/Pacific Islander students (10%). Most of the students had grade point averages in the C's and D's and were characterized by their teacher as "middle to low" achievers. Quite a few were English-as-a-Second Language students, but the teacher did not provide exact numbers. These second pilot classes were part of a special "Computer Academy" program for students who are at risk of not graduating due to their poor attendance. The entire cohorts of student participants in this program stay together from class to class, and there is one teacher per subject.

*IVC. Findings from the Climate Assessment Pilot Tests*

*Method.* The DIGS research team took a sample of student responses from the assessment pilot testing, then scored them, using item-by-item analytic rubrics that the team developed in the process. Qualified scorers were then brought in to use the rubrics to score all the student assessment responses, following a training process in which the scorers were presented with examples of student responses at each score point per item then given an opportunity to practice the rubrics with a common set of pre-scored training papers. There were in addition calibration papers. Inter-rater reliability statistics were compiled from the 414 responses that both raters scored from the training papers, calibration papers, and double-scored papers.[4] The raters discussed all responses that they double-scored so that they could find disagreements and reach consensus. Of all the student responses that were double-scored, the inter-rater reliability agreement was 91.1%. Of the 8.9% of disagreements, most were quickly resolved through discussion. 0.9% of the disagreements were substantive, which meant that the disagreements were indicative of problems with the rubrics. Because all of the disagreements were verbalized during discussions in the training sessions, the rubrics could be revised and training paper responses revisited before the scorers commenced with the scoring of the main body of responses to the focal item. The most significant rubric revisions involved changing the number of scale points to facilitate more consistent interpretation.

*Assessment results from the Climate module pilot test.* To permit the ranking of mean student performance per item, regardless of the size of the scale, the researchers converted the means to a common 0-1 metric (p values) by dividing the mean of the scores for the item into the number of scale points in the item.

*Results.* Table 1 displays the aggregate p values by aligned National science standards aligned to the items (1996) sorted from highest to lowest value.[5]

| **Standard** | **Number of items** | **P value (mean on 0-1 metric)** |
|---|---|---|
| Use technologies to collect, organize, and display data | 1 | 0.87 |
| Radiation of heat | 2 | 0.77 |
| Human-induced changes to atmosphere | 3 | 0.74 |
| Plan method | 1 | 0.73 |
| Review, summarize, and explain information and data | 14 | 0.70 |
| Formulate testable hypothesis | 2 | 0.63 |

[4] The numbers of responses used to calibrate these statistics varied slightly per item because the numbers of responses from the training set that were used as exemplars varied. When the researchers noticed in their pre-scoring of the training set that the responses to a particular item were yielding a wide diversity of responses, the researchers assigned more of the responses to the exemplars in the scoring guide and less to the calibration set. The converse was true if the researchers noticed less diversity in the responses.
[5] The numbers of items per standard were determined according to what sorts of interdependent sub-tasks would be most appropriate to the overall goals of the broad assessment task, not to ensuring equivalent distributions of items per standard.

| | | |
|---|---|---|
| Logical connections between hypothesis and design | 1 | 0.58 |
| Construct a reasoned argument | 2 | 0.54 |
| Scientific skepticism | 1 | 0.48 |
| Critique explanations according to scientific understanding, weighing the evidence, and examining the logic | 1 | 0.36 |
| Interactions within and among systems result in change | 1 | 0.29 |

Table 1: Aggregated DIGS Climate Assessment results by National Science standard

The mean p value across the 29 items on the assessment was .66. The results from Table 1 show that the student responses to an item aligned to the standard about using technologies to collect, organize, or display data received the highest mean score (.87) whereas an item designed to assess understanding about how change results from interactions among systems got the lowest mean score (.29). The mean of the 14 items calling for making analytical observations from data got higher mean scores (.70) than most of the other inquiry items. Means of items requiring application of knowledge about radiation of heat (e.g., urban heat island effects) and human-induced changes to atmosphere were high (.77 and .74 respectively). Essentially, students performed best on items requiring use of technology to perform tasks with fairly low cognitive demand and on items that tap their understanding of content. Scores were lower on inquiry tasks that were not about data analysis. The lowest scores however were received on items that required critical thinking (construct a reasoned argument; scientific skepticism; and critique explanations according to scientific understanding, weighing the evidence, and examining the logic) and general systemic thinking.

The following were some of the culminating questions in the assessment that were asked of the students after they interpreted specific data sets showing different environmental factors with implications for climate change. In parentheses are the phases of civic reasoning (i.e., phases 3, 5, and 6) to which each question is aligned. In italics, for illustrative purposes, are examples of relatively high quality student responses, as differentiated by clarity, and completeness. In parentheses after the response is a short explanation of the assigned score.[6]

[6] More detail about the scoring methodology including all constructs and scoring trait information is contained in the scoring guide at http://digs.sri.com/Climate/teacher.html

1. Do the charts and maps of data about Chicago convince you that the climate in Chicago is getting warmer? If yes is your answer, what data in particular? If no is your answer, why not, and what other data should be collected to investigate further? (Phase 3)

*Chicago's temperature is getting warmer, but very slowly. The temperature change is not nearly as rapid as that of Phoenix. Chicago's population has decreased, but its Physical size has increased, however only by a small amount.*

(Reasonable analyses of three data sources)

2. To reduce urban heat island effects, the Chicago City Council may carry out one of the actions listed below. Select one that you recommend to them and explain your reasoning (Phase 5)

a. Build more highways that go through the city
b. Give tax breaks to people who construct lighter-colored roofs and pavements
c. Make people drive cars that burn less gasoline per gallon
d. Plant more trees along city streets
e. Require power plants to produce electricity through solar power instead through burning coal

*By giving people tax breaks on constructing lighter-colored roofs and pavements, it would help reduce the urban heat island effect. This is because too much heat is already trapped in the city, so using lighter colors will retract the heat back into the air*. (The student's selection is grounded in an understanding of the causes of urban heat island effects)

3. The Chicago City Council does what you recommend to reduce urban heat island effects. A few years later, the Council wants you to evaluate if the policy is successful. Below is a list of possible types of data you might collect. Select two types you think should be collected. Explain your reasoning. (Phase 6)

a. carbon emissions from cars
b. maximum temperature
c. minimum temperature
d. ozone
e. rainfall
f. wind

*I would have them measure the mean temperature, because that provides a general idea about whether there has been any climate change overall, and I would have them measure the humidity, because water acts as a green house gas and therefore humidity should be taken into account when trying to measure the affect of carbon pollution as well as the heat island affect.*(Argues adequately for why it would be appropriate to collect mean temperature and humidity data)

4. The City Council has given you a limited budget for collecting data. How often do you think the data should be collected? Select one answer, then explain your reasoning. (Phase 6)

a. once a day for an entire summer
b. once a week for a year

c. once a month for five years
*I select "once a month for five years" because you have to allow enough time for change to occur and you also have to take into account natural seasonal and annual variation.* (Shows understanding of the need to differentiate between short-term weather variability and long-term trends when seeking evidence of climate change)

5. Imagine that ten years have gone by since the City Council passed a policy to reduce how much carbon people emit in Chicago. However, temperatures have not gone down. What might explain why? (Phase 6)
*The lack of monitoring temperatures every day over five years may relate to a lack of trend. Not as many people may have maid their roofs or pavements lighter, Chicago may have built taller buildings, the population may have increased, people may not have adopted carpool systems or used public transportation, and people may have driven more.* (Providing a wide range of possible explanations for why the intended effects of the policy may not end up being evident)

## V. CONCLUDING COMMENTS

In this paper we argued that educational design based on a full model of civic reasoning will enhance broad understanding of scientific, social, economic, and cultural aspects of important world issues and, in addition, may increase students' abilities to engage more productively and reflectively in public discourse. We offered examples from our own work of how the phases of civic reasoning can be employed to inform the design of science instruction and assessment.

Our first case was from an online college course in basic cognitive and neuroscience. Two lessons offered early in the course – one on argumentation and one on hypothesis testing -- introduced students to normative models for good thinking. The goals of these lessons were to enhance students’ future collaborative learning of science from the course and beyond. The lesson on argumentation was grounded in a collaborative problem emphasizing phases 1 (research problem evaluation), 2 (leveraging funding) and 5 (taking informed action) within the civic reasoning model. The hypothesis-testing lesson was framed by a collaborative problem that focused on phases 3 (research), 4 (peer critique) and 6 (evaluation design). Examination of student discourse showed that the two lessons evoked different forms of reasoning and attentional focus that could be explained in terms of the phases of civic reasoning. Students in the argumentation lesson examined values and reasoning related to selecting a problem for research funding, and they unexpectedly deliberated about the nature and quality of narrative versus scientific evidence, an important distinction. Students in the hypothesis-testing lesson addressed practical and technical issues of research design and grappled with how to operationalize variables and create measures, but they did not address questions regarding value of the research problem or the nature of evidence. Consideration of what is gained and what is omitted from each lesson serves to emphasize the importance of exposing students to tasks that engage them in all phases of civic reasoning.

The second case was based on an implementation of Practicing Democracy through Deliberative Discourse (PD3), an adaptable instructional approach that focuses on phase 1 of the civic reasoning model. PD3 illustrates how tasks can be designed to engage students in deliberative discourse to increase their metacognitive awareness of issues that come into play when attempting to seek common ground and focus for scientific research related to controversial issues. Based on implementation in both a teacher professional development and an 11th grade high school setting, this case illustrates how learners can be scaffolded to a better understanding of how their differences and bases for common ground are relative, depending upon the issue being addressed and it relatedness to other concerns such as self-interest, moral implications, and conduciveness to rational analysis.

Our third case came from Data Sets and Inquiry in Geoscience Education, an NSF online curriculum project that engages high school students in solving ill-structured problems focuses on phase 3, 5 and 6 of civic reasoning (designing and carrying out scientific research, using findings to set policy or take action, and evaluating policy effectiveness). The examples provided in this paper focus on assessment, demonstrating how questions are differently formulated to address the different phases. Examples of students' responses illustrate how targeting the different phases in assessment require different forms of student reasoning.

These cases were shared to illustrate how a full model of civic reasoning can inform and enrich the design of science instruction and assessment. We acknowledge that there are already many innovative attempts to engage in distributed knowledge building around science issues such as the Web-based Inquiry Science Environment (http://wise.berkeley.edu/) as well as virtual communities in which leaders and peers share knowledge and practice discourse. Yet, the extent to which current curricula, including these innovative projects, catalyze, or can be expanded to catalyze, attention to the full pantheon of civic reasoning remains to be determined and is an important question for future research.

Our examples provide proof of feasibility, but this must be followed by critical examination of how well programs based on a full model of civic reasoning can prepare young "digital natives" to employ their technological prowess constructively to more effectively engage in civic issues that have scientific import. Harnessing distributed intelligence in the service of solving critical world problems is a complex challenge that educational systems should address. We hope our modest proposal spurs movement in that direction.

We conclude by offering up some research questions about what characterizes high-quality learning and instruction in the six phases of civic reasoning that might be addressed using qualitative and quantitative methods through instructional research within a design-based research (DBR) (Collins, Joseph, & Bielaczyc, 2004) or in vivo experimentation (e.g., Aleven & Koedinger, 2002) paradigms.

- How do we best identify and select controversial topics to support a science curriculum based on civic reasoning, giving consideration to issues such students' agency and motivation, availability of supporting materials, importance for community and human welfare, and a topic's conduciveness to generating in-depth discourse about and understanding of fundamental science principles?

- What alternative instructional designs (e.g., forms of question-framing about individual issues, numbers of issues, discourse and response paradigms) are most conducive to the best possible outcomes, and how can we systematically evolve and share them widely with practitioners?
- To what extent does the success of P3D or other like-minded approaches depend on students' prior ability to construct and recognize valid arguments, and must we prepare students in advance to learn effectively through argumentative discourse (Cavagnetto, 2010)?
- What alternative feasible forms of scaffolding (e.g., examples of how to respond to different stimuli, step-by-step directions, coaching, peer learning) are most conducive to the best possible outcomes, giving consideration to such issues as depth of teachers' understanding and capabilities in a world of increasing class sizes and decreasing resources for education?
- How do different types of students (e.g. science majors, non-science majors, low achievers, high achievers, students with learning disabilities, etc.) respond to civic reasoning approaches generally and to different implementation methods specifically?
- What different forms of learning and reasoning are promoted by instructional interventions that focus on different phases of the civic reasoning model?
- How do cognitive outcomes of science instruction based on the civic reasoning model interact with affective outcomes, and how do these interactions influence students' knowledge of and interest in science as both citizens and as potential science professionals?
- Which instructional components evidenced to work well in face-to-face situations can be equally successful in online situations, and vice versa?
- How can we best leverage cyberinfrastructure to promote a civic reasoning approach to science instruction on a large scale?

If we agree that the demands of 21st-century society require the implementation for all of a "thinking curriculum" (Resnick, 2010), answers to these research questions are worth pursuing. We need to know how school curricula can be widened to prepare students for the six stages, how we can get all students to know enough about science to contribute to the civic engagement process in the most informed and astute manner, and ascertain to what extent those who go on to pursuing science careers end up performing better in those careers if they are better able to gauge their own technical contributions within context of the full pantheon of civic reasoning outlined here.

CITATIONS

Abbeduto, L. & Symons, F., Eds (2010). Clashing views in educational psychology ($6^{th}$ Ed.) New York: McGraw-Hill.

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26(2).

Bardach, E. (1977). *The implementation game: What happens after a bill becomes a law*. Cambridge, MA: The MIT Press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education , 24*, 61-101. Washington DC: American Educational Research Association.

Cavagnetto, A.R. (2010). Argument to foster scientific literacy: A review of argument interventions in K-12 science contexts. *Review of Educational Research*, 80(3), 336-371.

Coles, G. (2004). Danger in the classroom: “Brain glitch” research and learning to read. *Phi Delta Kappan*, January, 344-351.

Cohen, Michael D., James G. March, Johan P. Olsen A Garbage Can Model of Organizational Choice (1972). Administrative Science Quarterly, Vol. 17, No. 1. pp. 1-25.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, 13(1), 15-42.

Derry, S. J., Hmelo-Silver, C. E., Nagarajan, A., Chernobilsky, E., & Beitzel, B. (2006). Cognitive transfer revisited: Can we exploit new media to solve old problems on a large scale. *Journal of Educational Computing Research*, 35, 145-162.

Eagan, B. R. (2010). Unpublished Master’s Thesis. Department of Educational Psychology, University of Wisconsin-Madison.

Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in the classroom contexts. *Review of Research in Education, 30, 1-32.*

Freire, P. (1995). *Pedagogy of the oppressed.* New York: Continuum.(Original work published 1970)

Gruenewald, D. A. (2003). The best of both worlds: A critical pedagogy of place. *Educational Researcher 32(4), 183-197.*

Halpern, D. (2003). *Thought and knowledge* (4th ed.). Mahwah, NJ, Erlbaum.

Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.

Hirsch, E. D. Jr. (1996). Reality’s revenge: Research and ideology. *American Educator,* Fall, 4-6; 31-46

Hutchins E. (2001). Distributed cognition. In N. J. Smelser, J. Wright & P. B. Baltes (Eds.), *The international encyclopedia of the social and behavioral sciences* (pp. 2068-2072). Elsevier.

Jensen, E. P. (2008). A fresh look at brain-based education. *Phi Delta Kappan,* February, 409-413; 414-417.

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

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

March, J. G. & Olsen, J. P. (1979). Ambiguity and choice in organizations. 2nd edition, Bergen: Universitetsforlaget [LB2806.M353.1979]

National Center for Education Statistics (2011). *Common Core of Data.* Retrieved from http://nces.ed.gov/ccd/

*National science education standards*. (1996). Washington, D.C.: National Research Council/National Academy Press.

Resnick, L (2010). Nested learning systems for the thinking curriculum. *Educational Researcher*.April, 39(3), 183-197

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

Quellmalz, E. S., & Zalles, D. R. (2010). Datasets for inquiry in geoscience: A design model. *Int. J. Learning Technology, 5(4.)*

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky. Volume 1: Problems of general psychology*. New York: Plenum (edited and translated by N. Minick)

Wertsch, J. V. & Kazak, S. (2011) Saying more than you know in instructional settings. In T. Koschmann (Ed), *Theories of learning and studies of instructional practice* (pp. 153-166) *Springer*

Windschitl, M. (1999). The challenges of sustaining a constructivist classroom culture. *Phi Delta Kappan*, June, 751-755.

Zalles, D., Quellmalz, E., Gobert, J., & Pallant, A. (2007). *Data Sets and Inquiry in Geoscience Education: Final Report*: SRI International.