Argument Focused Instruction and Science Proficiency in Middle and High School Classrooms

Patrick J. Enderle
The Florida State University

Jonathon A. Grooms
The Florida State University

Victor D. Sampson
The Florida State University

Symposium Paper presented at the 2012 NARST Annual International Conference
Indianapolis, IN
March 28, 2012

Introduction & Background

Science teachers in the United States tend to provide their students with several different laboratory experiences[1] over the course of a school year to help them understand the content outlined in a state or district curriculum. These laboratory experiences tend to be designed and implemented in the same manner regardless of grade level or course. Most science teachers first introduce their students to a concept and then give them a worksheet taken from a lab manual that includes a procedure to follow, a data table to fill out, and a set of analysis questions to answer with the hope that the laboratory experience will illustrate, confirm, or otherwise verify the concept that was introduced to students at the beginning of the activity. This type of approach, however, is an ineffective way to help students understand the content under investigation, learn how to engage in important scientific practices (such as investigation design, argumentation, and writing) or develop scientific habits of mind (Duschl et al., 2007; National Research Council, 2005).

As a result, most laboratory experiences in United States do little to promote and support the development of science proficiency. There are three main reasons for this. First, this type of laboratory-based instruction is not well aligned with the design principles for developing an effective laboratory experience for students that have been outlined by the National Research Council (2005). Second, this type of instruction does not reflect current findings from research about how people learn (Abell & Lederman, 2007; Bransford, Brown, Cocking, Donovan, & Pellegrino, 2000; National Research Council, 2008). Third, and perhaps most importantly, this type of traditional approach to laboratory-based instruction was never designed to promote and support the development of science proficiency.

Scientific proficiency, as described by Duschl, Schweingruber, & Shouse (2007), encompasses a variety of knowledge and skills required by an individual to function effectively in an increasingly complex, information-driven society. The framework of scientific proficiency describes science as “both a body of knowledge and an evidence-based, model-building enterprise that continually extends, refines, and revises knowledge” (p. 2). In this view, individuals that are proficient in science: (a) know, use, and can interpret scientific explanations of the natural world; (b) can generate and evaluate scientific explanations and arguments; (c) understand the nature and development of scientific knowledge; and (d) can participate in the practices and discourse of the various scientific disciplines in a productive manner.

By implementing instructional strategies that focus on scientific proficiency, classroom instruction shifts from traditional, prescriptive activities to those that afford students the

---

[1] We define a laboratory experience as “an opportunity for students to interact directly with the material world using the tools, data collection techniques, models, and theories of science” (NRC, 2005, p. 3).

opportunity to engage in the practices and discourse of science (Duschl et al., 2007; National Research Council, 2005, 2008). The Argument-Driven Inquiry (ADI) instructional model (Sampson, Grooms, & Walker, 2011) is one strategy that is designed to foster the development of the four key aspects of scientific proficiency. Classroom activities structured according to the ADI model engage students in data collection and analysis, argument generation, group argumentation, scientific writing, and double blind peer review processes. The ADI instructional model therefore is well aligned with various aspects of the scientific proficiency framework and provides a way for students to develop the knowledge and skills they need to be proficient in science while in school. Figure 1 provides a representation of the several component stages of the ADI model.

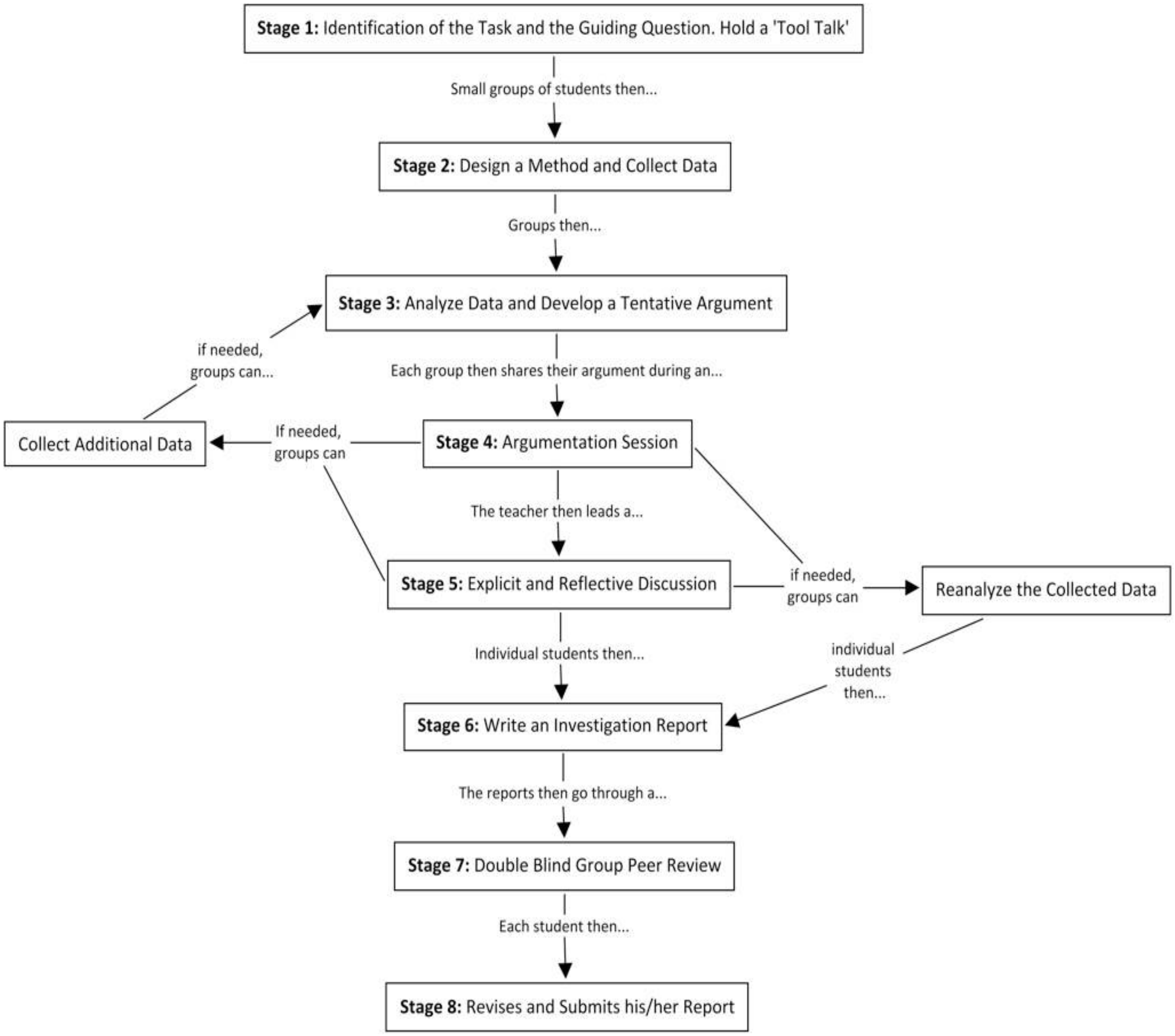

Figure 1: Stages of the ADI Instructional Model

This study stems from a yearlong implementation of the *Argument Driven Inquiry* (ADI) instructional model (Sampson, Grooms, & Walker, 2011) in middle and high school classrooms. This instructional model provides an alternative approach to designing and conducting classroom laboratory activities. **Specifically, this study concerns the measurable changes demonstrated in students' abilities to understand relevant science content and correctly apply that knowledge to explain unique scenarios**. **Furthermore, the study also explored middle and high school students' learning to write scientifically following the year of ADI instruction.** The design of the ADI instructional model is based on a hypothesis that efforts to improve science proficiency will require the development and continued use of laboratory experiences that are more authentic and educative. Figure 2 provides a graphical description of this working hypothesis.

Figure 2: Hypothesis Describing the Potential Impact of Implementing ADI Instruction

Context of the Study

The study described here occurred during year one of a larger, three-year project aimed at assessing students' improvements in science proficiency as a result of experiencing argument focused instruction. The research setting involved the high school biology (2 teachers; 8 sections) and chemistry (2 teachers; 8 sections) courses and the middle school life science (1 teacher; 5 sections) and physical science (1 teacher; 4 sections) courses at a university research K12 school. Each class engaged in at least 8 unique argumentation investigations over the course of the 2010-2011 school year, with some completing as many as 14. These investigations, which were designed by the researchers and project teachers, focused on several major themes in the courses, such as cell theory, evolution, exothermic and endothermic reactions, Newton's 2nd law, and several others.

Methods

An assortment of assessments was given to the students at the beginning and end of the year. The assessments used in this study focus on the key aspects of scientific proficiency. The assessments were designed to align with the major focal concepts for the argument based instructional activities in each course. All assessments were scored using rubrics developed by the research team and scientific experts. A pair of research team members scored at least 25%

of the full set of each assessment, which had been blinded concerning student identity and pre/post timing. The intraclass correlation coefficient (ICC), a measure of reliability similar to Cohen's Kappa and interpreted using the same scale, was determined for each team (two-way random effects, absolute agreement). An ICC above 0.6 is considered substantial agreement (Landis & Koch, 1977), and once this level of agreement was determined, the team members scored the remainder of their assessment sets individually.

Content Knowledge Assessment

To measure changes in students' understanding and use of science content, subject-specific content assessments were developed by the research team and project teachers. Each assessment is comprised of eight free response questions, each related to one of several "Big Topics" in the related subject area. Each question includes an opening paragraph that provides a relevant scenario or context, followed by two questions. The first question asks the student to *describe* the fundamental science concept (*Know*) and the other asks the student to *apply* that concept to explain the scenario provided (*Use*). The rubrics for these assessments were developed from answers provided for the questions by expert scientists who have experience with K12 science education. A students' score was developed from the rubric based on correct description of several content elements identified in the expert's answer to the question. The rubrics were then scaled so that individual questions could be compared. The scoring teams for this assessment achieved ICCs of 0.897 (Biology), 0.900 (Chemistry), 0.986 (Life Science), and 0.970 (Physical Science).

As a point of comparison for these assessments, a more "traditional" assessment was constructed using released multiple choice questions from several prominent standardized tests used as benchmark measures of student learning, including the NAEP, PISA, and TIMMS. These questions resemble the typical measures for gains in content knowledge that are ubiquitous in K12 education.

Scientific Writing Assessment

Another aspect associated with the science proficiency framework involves students' abilities to develop and evaluate scientific explanations and arguments. In the context of the study presented above, this facet of science proficiency was measured using an individual writing assessment. This assessment provides a student with a small amount of background information and a related data table followed by a prompt. The prompt presents an argument by a scientist/expert who provides an inaccurate explanation for the data. The students are directed to generate their own argument in response to the scientist's claim by arguing in support of a countering claim, including evidence and a rationale based on the data and information provided in the question. Structuring the writing assessment in this manner allows the student an opportunity to evaluate and critique the scientific argument of the expert and then develop an argument for a counter claim.

The research team developed a general rubric for scoring the writing assessment, which had a maximum score of 28 points and was divided into three subscales: *Argument Structure, which* focuses on the inclusion of fundamental argument components (i.e. claims, evidence, and rationale) (6 points); *Argument Content*, which concerns the quality and relevance of the argument components with respect to the specific task/prompt (10 points); and *Mechanics*, which addresses the punctuation, grammar, and technical quality of the writing (12 points). Four teams of two researchers each scored the sets of writing assessments for Chemistry, Biology, Life Science, and Physical Science, including the pre-, mid-, and post-intervention versions. Each team scored at least 10% of the set together in order to calculate inter-rater reliability. The ICCs (two-way random effects, absolute agreement) for the different teams were: 0.71 (biology), 0.76 (life science), 0.74 (physical science), and 0.62 (chemistry). These scores demonstrate substantial agreement between the raters (Landis & Koch, 1977); however, the ICC for Chemistry was lower than desired, therefore a consensus score was generated for each student, the consensus scores were used in all analyses. The remainder of the assessment set for biology, life science, and physical science was divided between the raters and scored individually. These data were then analyzed using repeated measures general linear modeling.

Analysis & Results

Content Knowledge Assessment

A Paired Samples t Test analysis of student responses on the subject specific content assessments demonstrated a significant difference between the pre and post administrations. The effect size (as measured by Cohen's *d*) of all of these comparisons were large (>0.8) in magnitude, as seen in Table 1. Students' overall scores on the content knowledge assessment increased significantly in every course, with Biology producing the largest effect. The sub-scale comparisons also demonstrate a significant difference in scores from pre to post assessment administration. The effect sizes for each comparison support the strength of the changes observed in the scores, offering evidence that students became more proficient in knowing and using scientific explanations after engaging in a year of classroom laboratory instruction designed using the ADI model.

Table 1: Paired Samples t Test Results for Subject-Specific Content Assessments

Scale/Subject	**Pre Mean**	**Pre St. Dev.**	**Post Mean**	**Post St. Dev.**	**Change**	**t**	**df**	**Sig.**	**Cohen's *d***
Total Score									
Biology	4.45	3.74	17.89	6.51	13.44	21.78	110	p<.001	2.53
Chemistry	2.34	2.060	9.07	4.919	6.726	15.010	105	p<.001	1.78
Physical Sci	3.78	2.545	8.08	3.543	4.300	9.764	59	p<.001	1.39
Life Sci	4.50	3.160	11.84	5.730	7.342	11.842	75	p<.001	1.57
Know Subscale									
Biology	1.57	1.64	8.93	3.83	7.36	20.77	110	p<.001	2.50
Chemistry	1.46	1.429	5.10	3.049	3.642	12.789	105	p<.001	1.53
Physical Sci	1.70	1.522	3.92	1.749	2.217	9.054	59	p<.001	1.35
Life Sci	2.22	1.887	6.43	3.087	4.211	11.466	75	p<.001	1.65
Use Subscale									
Biology	2.90	2.54	8.96	3.24	6.06	20.81	110	p<.001	2.08
Chemistry	0.88	0.992	3.96	2.418	3.085	13.874	105	p<.001	1.67
Physical Sci	2.08	1.488	4.17	2.505	2.083	6.256	59	p<.001	1.01
Life Sci	2.28	1.654	5.41	3.142	3.132	8.664	75	p<.001	1.25

To further understand students' development of their proficiency in understanding and using scientific explanations, normalized learning gains (NLGs) were calculated overall and for both the *Know* and *Use* subscales. Normalized learning gains are calculated through subtracting a student's score on the pre-test from both their post-test score and the total possible score and dividing these numbers (i.e. – (Post – Pre)/(Total – Pre) = NLG). The advantage of using NLGs for studying learning involves the accounting of students' prior knowledge inherent in the measure. Thus, NLGs provide the ability to track growth that occurred during the intervention period, as opposed to typical assessments that only provide a one-time snapshot of learning. The NLGs from the content knowledge assessment and related subscales were compared to similar measures for the students' responses on the standardized question assessment. Figure 3 visualizes this comparison. The NLGs observed for each course were relatively consistent when comparing overall score and subscale scores. Another interesting result was the trend that in most courses, the students actually scored better on the content knowledge assessment where they were asked to write out their thoughts when compared to the standardized question assessment, which involves only multiple choice questions.

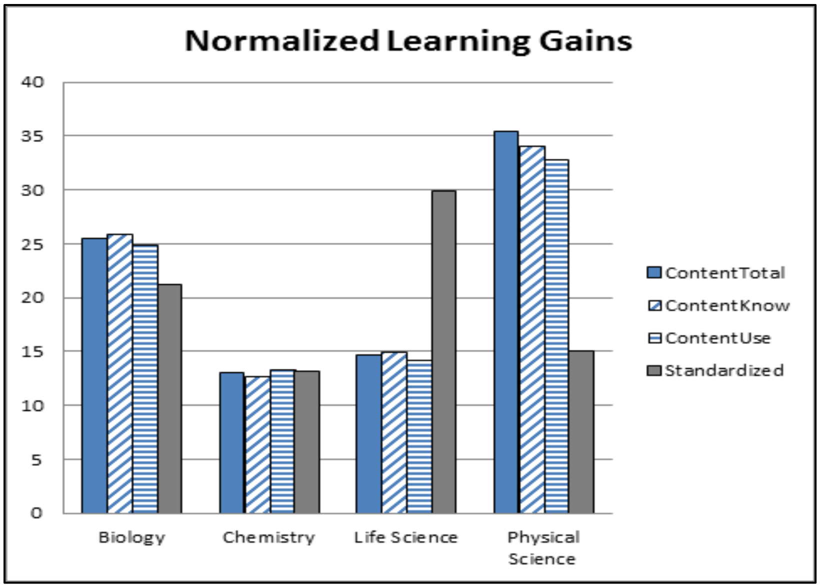

Figure 3: Normalized Learning Gains Comparison

Scientific Writing Assessment

Figure 2 displays the trend in students' scores for the scientific writing assessment over the course of the school year. Students' writing scores increased significantly during the three administrations of this assessment. A consistent, linear increase is seen in the Biology and Life Science overall scores during the course of the year. However, writing scores for students in the Chemistry and Physical Science courses follow a noticeable plateauing trend in the second half of the year, which will be discussed later. Table 5 further illustrates the significance of each change from the pre-intervention assessment to the post-intervention assessment for each course, as determined by repeated measures ANOVA, for the overall writing score and specific sub-scales. The sub-scale comparisons offer evidence that students' proficiencies in writing and generating scientific arguments improved over the course of a year where they experienced ADI instruction.

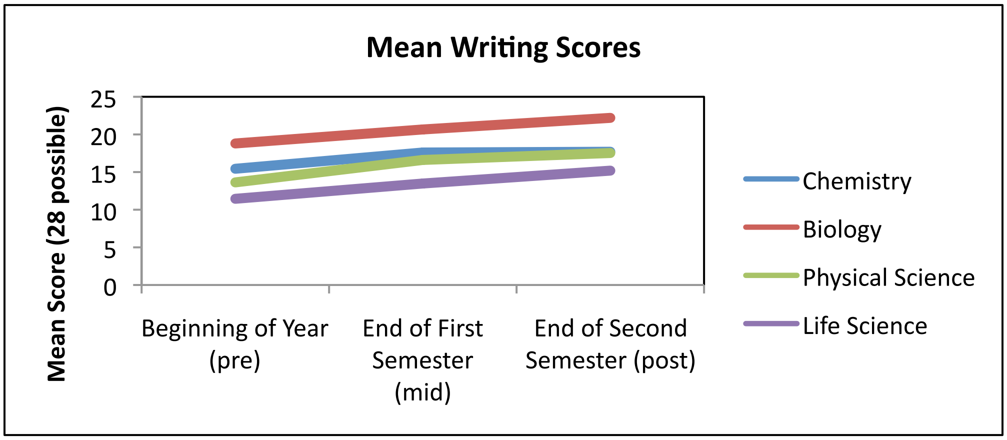

Figure 4: Mean writing scores for each course across the school year

Table 2: Writing assessment repeated measures ANOVA results; Total and by subscale

	Biology	Chemistry	Life Science	Physical Science
Total Score	F(2) = 45.31, *p* < .001	F(2) = 20.32, *p* < .001	F(2) = 16.92, *p* < .001	F(2) = 23.01, *p* < .001
Argument Structure	F(2) = 28.76, *p* < .001	F(2) = 38.34, *p* < .001	F(2) = 17.36, *p* < .001	F(2) = 9.64, *p* < .001
Argument Content	F(2) = 44.92, *p* < .001	F(2) = 16.14, *p* < .001	F(2) = 6.31, *p* = .002	F(2) = 16.81, *p* < .001
Writing Mechanics	F(1.91) = 6.69, *p* < .01*	F(2) = 1.34, *p* = .57	F(2) = 6.63, *p* = .002	F(2) = 12.59, *p* < .001

*Greenhouse-Geisser correction was used due to a violation of the sphericity assumption

Findings

ADI Impact on Students' Content Knowledge

After a year of argument focused instruction, middle and high school students demonstrated significant learning gains in their understanding and ability to apply science content knowledge. The Biology and Physical Science courses demonstrated the most gains. The Biology course completed 14 different ADI investigations. The Physical Science course completed 8 ADI investigations. Although comparing the *Know* and *Use* subscales in each course indicates that students' ability to describe the relevant scientific concepts was slightly better in most cases, their relatively similar ability to apply that knowledge is encouraging. The large effect sizes for each of the comparisons, including the subscale comparisons, demonstrate that students' learning in regards to content knowledge was substantial.

A comparison of the subject specific content assessment results to a collection of standardized content questions demonstrates that for most subject areas, students demonstrated similar, if not more, learning on the more in-depth assessment. The reverse of this result seen in the Life Science course could possibly be due to teacher effects, as this teacher was in her first year of alternative certification. This finding offers support for the study hypothesis that using argument focused instruction, such as the ADI model, can provide a richer learning environment that increases students' science proficiencies. At minimum, the implementation of ADI activities, which require more class time due to several stages of activities, in place of more traditional, verification labs does not hinder a student's learning of content knowledge.

ADI Impact of Students' Scientific Writing

After a year of ADI instruction, middle and high school students demonstrated significant gains in their abilities to generate and evaluate scientific explanations and arguments, as measured by the scientific writing assessment. The content and structure of their counter-arguments improved as well as the general mechanics associated with their writing skills (chemistry excluded). These findings demonstrate that the multiple components of the ADI instructional model can help students develop their scientific proficiencies in other areas beyond content. Students' ability to communicate science through writing and their ability to generate and evaluate scientific arguments both increased after a year of instruction where classroom laboratory experiences followed the ADI instructional model. These findings offer further support for the study's original hypothesis that implementation of science labs following the ADI model can help to improve students' science proficiency.

The writing scores of the chemistry and physical science students seemed to plateau between the mid-year assessment and end of the year assessment. Both of these courses were heavily affected by the annual state standardized assessment, wherein each course experienced a reduced number of argument focused lessons, which may have served to diminish the potential gains during the second semester of the school year. In contrast, the Life Science and Biology courses were able to complete 12 and 14 ADI investigations, respectively. This finding indicates a potential dosage effect in regards to the amount of improvement in students' science proficiency. That is, the amount of writing the students were engaged in was less for the students in the Physical Science and Chemistry courses. It is plausible that this decreased engagement in the writing of investigation reports, a critical element of the ADI model, contributed to the plateau effect seen in these scores. Also, decreased exposure to argument generation and evaluation activities inherent in the ADI model could have contributed to the decrease in overall writing scores, although pre to post differences were still significant.

Implications of the Study

The results of this study point to some broader implications for science education, most specifically in K12 settings. The ADI instructional model demonstrates potential for serving as an approach to designing classroom science laboratory activities that are more meaningful than more traditional laboratory instruction found in many classrooms (NRC, 2005). The variety of activities present throughout the various stages of this model engages students in different tasks that require the employment of several unique proficiencies. In the ADI model, students are challenged to generate their own investigation design and develop their own scientific argument in response to a guiding research question, where they must justify their claim through knowing and using major scientific concepts. Students' continue to develop their proficiency in argument generation and writing scientifically as they create investigation reports and participate in peer reviews of their classmates. Thus, argument based instruction, such as

ADI, involving more authentic scientific practices can benefit students' learning of science in multiple ways, helping to develop science proficiency and moving beyond a sole focus on mastery of content knowledge.

Furthermore, the emerging Common Core State Standards for Science (NRC, 2011), developed and supported by a majority of states, move beyond a primary concern for content knowledge to encompass performance expectations that reflect ideas described in the science proficiency framework. Typical high-stakes standardized assessments are limited in that they do not offer a complete picture of a student's science proficiency. The assessments employed in this study provided a more meaningful insight into students' capabilities when compared to the standardized assessments that are the norm. As science education as a whole moves toward a broader framework that encompasses goals related to multiple science proficiencies, assessments must also develop so they can provide teachers and educators with a more meaningful understanding of their students' learning. The two assessments described in this study offer potential models for assessments aligned with frameworks encompassing multiple proficiencies. Thus, as standards and assessments progress to include the aspects of science proficiency, so to must science teaching address the development of this multifaceted construct. Again, classroom instruction that engages students in more authentic scientific practices, such as the stages of activity in the ADI model, have the potential to assist teachers in broadening their teaching and developing multiple proficiencies in their students.

Limitations

The authors readily acknowledge the results and findings presented here are limited in their impact due to limitations in the design of the study. First, this study is of an exploratory nature, and as such, did not distinguish a comparison group to determine how much science learning occurs in the absence of ADI instruction as measured by these assessments. The purpose of this initial study was to explore whether or not ADI instruction had any impact, positive or negative, on the development of science proficiency. Since the only laboratory instruction these students received during this year was through using the ADI model, the growth in their science proficiencies can be correlated to the effects of ADI instruction. To further determine causal relationships and their magnitude, future studies will need control groups identified and assessed to make any such generalizations. Currently, these studies are being conducted.

Another limitation to the study presented here is the lack of inclusion of any teacher factors. Although all the teachers that took part in this study were trained in the same manner, the fidelity of implementation of the ADI model varied in their individual classrooms. Even teachers within the same discipline subject area can enact vastly different versions of any instructional or curricular model (Smith & Southerland, 2007). Instruments measuring this aspect are in development and data related to teacher factors is being collected in the current studies to account for such variations

The research reported here was supported by the Institute of Education Science, U.S. Department of Education through grant #R305A100909 to Florida State University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

References

Abell, S., & Lederman, N. (2007). *Handbook of research on science education.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bransford, J., Brown, A., Cocking, R., Donovan, S., & Pellegrino, J. (Eds.). (2000). *How people learn: Brain, mind, experience, and school (Expanded Edition).* Washington, D.C.: National Academy Press.

Duschl, R., Schweingruber, H., & Shouse, A. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

National Research Council. (2005). *America's Lab Report: Investigations in High School Science*. Washington D.C.: National Academy Press.

National Research Council. (2008). *Ready, Set, Science: Putting Research to Work in K-8 Science Classrooms*. Washington, D.C.: National Academy Press.

Sampson, V., Grooms, J., & Walker, J. (2011). Argument-Driven Inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study. *Science Education, 95*(2), 217-257.

Smith, L., & Southerland, S. (2007). Reforming practice or modifying reforms?: Elementary teachers' response to the tools of reform. *Journal of Research in Science Teaching, 44*(3), p. 396-423.