THE DEVELOPMENT OF SCIENCE PROFICIENCY IN HIGH SCHOOL CHEMISTRY STUDENTS ENGAGED IN ARGUMENT FOCUSED INSTRUCTION

Patrick J. Enderle
The Florida State University

Jonathon A. Grooms
The Florida State University

Keisha Williams
Florida State University Schools

A paper presented at the 2012 Annual Meeting for the American Educational Research Association in Vancouver, British Columbia, Canada
April 16, 2012

OBJECTIVE

This study explored the development of several aspects of science proficiency in high school chemistry students that experienced a year of argument focused laboratory instruction. The ADI instructional model (Author, 2011) shifts classroom laboratory investigations towards more productive interactions among students that engage several aspects of science proficiency.

THEORETICAL FRAMEWORK

Scientific Argumentation and Argument-Driven Inquiry

Argumentation in science represents "a logical discourse whose goal is to tease out the relationship between ideas and evidence" (Duschl, Schweingruber, & Shouse, 2007, p. 33). Discursive activity characterized as scientific argumentation involves the construction of knowledge claims supported through genuine evidence drawn from authentic inquiry and justification for those claims and evidence through connection to ideas and models privileged and accepted by the broader scientific community. The process of argumentation encompasses interactions in which individuals propose, support, critique, and refine ideas for the purpose of understanding the natural world (Driver, Newton, & Osborne, 2000; Kuhn, 1993; Sampson & Clark, 2011). These types of interactions become fundamental to the creation and evaluation of scientific knowledge, practices which serve to uniquely distinguish science from other ways of making sense of the world (Duschl & Osborne, 2002).

As central as argumentation is to the scientific enterprise, students in science classrooms are rarely afforded the opportunity to engage in these aspects of scientific practice, much less learn the epistemological commitments and warrants that separate scientific argumentation apart from other forms of argumentation (Duschl et al., 2007; National Research Council, 2005, 2008). Students must engage in authentic scientific practices in order to learn science, both concepts and skills, from their experiences. Rather than participate in laboratory experiences where they are provided with a predetermined set of procedures and organizations of data followed by several short analysis questions, students need opportunity to participate in the discursive practices of science, including the coordination of evidence and theory to support knowledge claims (Reiser, Tabak, Sandoval, Smith, Steinmuller, & Leone, 2001). Students must understand the practices, such as investigation design and collection of informative data, valued in science by experiencing them first hand. These experiences help students understand the types of methods that are privileged in science and more productive for generating scientific knowledge (Sandoval & Reiser, 2004).

As researchers and educators have come to understand the importance of scientific argumentation in the learning of science, several new instructional approaches and curricula have been developed to provide students more opportunities to learn about and how to meaningfully participate in this discursive activity. One such instructional model is called *Argument-Driven Inquiry* (ADI) (Sampson, Grooms, & Walker, 2011). The ADI model involves eight stages of educative activity that reflect the practices of science embedded within contexts that teachers can use to teach scientific concepts to their students. These stages engage students in the design of unique investigations for the purpose of answering a guiding research question through the generation of scientific arguments that are shared among their peers. During these design processes, students must work together to determine kinds of data that would be appropriate for answering the research question provided for them. Students must also decide on the appropriate methods they employ for collecting and analyzing data, engaging them in activities that model the development of scientific knowledge and understanding the underlying reasons for valuing certain approaches. The ADI model allows students to experience authentic investigations where they can meaningfully explore science content topics they are learning during other times in their science classes, applying them to unique scenarios. The stages also necessitate student participation in scientific discourse through writing expository and persuasive investigation reports and critiquing other students' writing and arguments through a blinded peer review process. Figure 1 provides a graphical representation of the ADI instructional model.

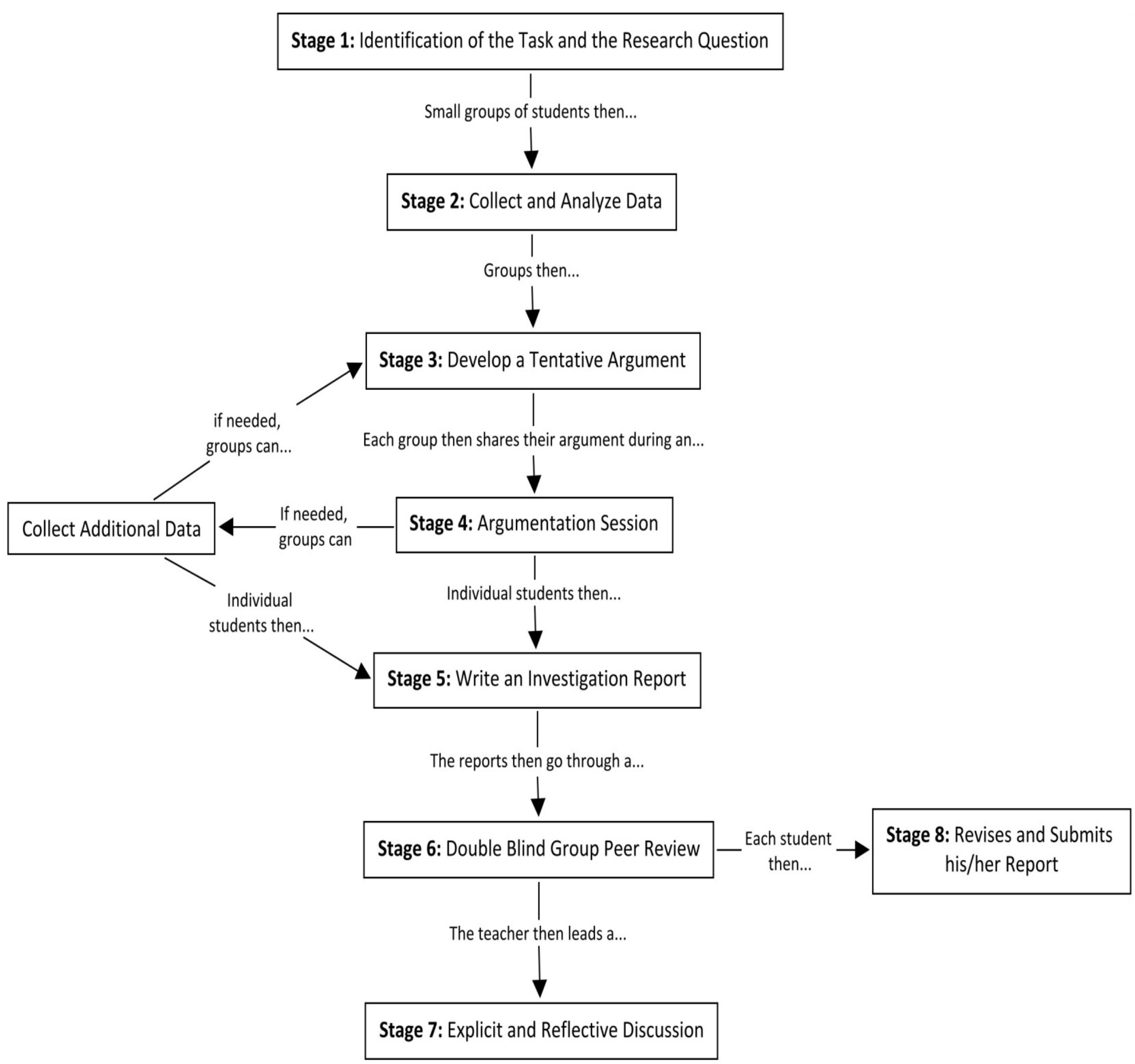

Figure 1: Stages of the Argument Driven Inquiry (ADI) instructional model

Science Proficiencies as Learning Goals for Science Education

Scientific proficiency has emerged as an updated and broader concept representing the fundamental science learning desired for K12 students, stemming from the ideas of science literacy that have been central to reform efforts of the past two decades. The multiplicity of meanings developed for science literacy (Roberts, 2007) necessitated that a more comprehensive construct embodying a variety of knowledge and skills be developed. Duschl, Schweingruber, and Shouse (2007) describe science proficiency to encompass a variety of knowledge and skills required by an individual to be able to function effectively in an increasingly complex, information-driven society.

The framework of scientific proficiency positions 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. The elements of science proficiency are also reflected in the language and substance of the emerging Common Core State Standards for Science (NRC, 2011), developed and supported by a majority of states, which move beyond a primary concern for content knowledge to encompass performance expectations and development of critical scientific practices.

By implementing instructional strategies that focus on scientific proficiency, classroom instruction shifts from traditional, prescriptive activities to those that afford students the opportunity to engage in the practices and discourse of science (Duschl, Schweingruber, & Shouse, 2007; National Research Council, 2005, 2008). The ADI instructional model 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 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.

The ADI instructional model most specifically targets the enhancement of laboratory experiences in science classrooms. The working hypothesis for this study predicts that students who engage in laboratory instruction designed using the ADI instructional model throughout the course of a school year will demonstrate improvement in several aspects of their science proficiency due to their engagement in more authentic scientific practices. In a broader sense, 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 offers a graphical representation of this hypothesis.

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

METHODOLOGY

The study described here occurred during year one of a larger, three-year project aimed at refining the ADI instructional model and assessing students' improvements in science proficiency as a result of experiencing ADI-based instruction (IES Grant #: R205A100909). The project is using an iterative outcome-focused approach that is consistent with the major tenets of design-based research (Brown, 1992; Brown & Campione, 1996; The Design-Based Research Collective, 2003) to develop and refine the ADI instructional model through several iterative cycles of design, enactment, analysis, and redesign. This research setting involved the high school chemistry courses at a university research K12 school.

Classroom Context

Two teachers were involved in teaching the chemistry courses where this study occurred. One of the teachers, also the third author of this paper, taught six sections of the chemistry course, with one of those sections considered to be "Honors". This teacher was in her first year of teaching and had recently completed a graduate degree focused on science education. The other teacher taught two sections of the chemistry course, both considered "Honors". Combined, these eight sections of chemistry course provided an initial sample size of 195 students. However, in light of parent consent and student assent requirements, the pool was further reduced by 35 students. Absenteeism on days of specific administrations of the assessments also affected the total sample size available for analysis for each assessment. Each class engaged in at least 12 unique ADI investigations over the course of the 2010-2011 school year. These investigations, designed by the researchers and project teachers, focused on several major themes in chemistry, including classification of matter, stoichiometry, chemical reactions, and several others.

Multifaceted constructs such as science proficiency require a variety of tools to assess students' knowledge and abilities related to science. The ability to know and use scientific content knowledge to solve and explain problems is a key component of scientific proficiency, which requires a unique assessment when compared to other aspects of scientific proficiency such as the ability to participate in the practices and discourse of a scientific discipline. Using one assessment to measure scientific proficiency would offer a biased view of students' abilities, as not all assessments are adequate for all learning outcomes. Thus, assessments aimed at understanding the development of science proficiency must be crafted with specific aspects in mind. Several assessments were given to the student subjects to complete. The assessments used in this study focus on the key aspects of scientific proficiency. Table 1 identifies the aspects of science proficiency and the accompanying assessment that will be addressed in this proposal.

Table 1: Aspects of Science Proficiency and Correlated Assessment

Aspect of Science Proficiency	Description	Assessment Instrument
Aspect 1	Students know, use, and can interpret scientific explanations of the natural world	Chemistry Content Knowledge Assessment
Aspect 2	Students can generate and evaluate scientific explanations and arguments	Chemistry Performance Task - Argument Generation Section
Aspect 3	Students understand the nature and development of scientific knowledge	SUSSI
Aspect 4	Students productively participate in the practices and discourse of the scientific community	Chemistry Performance Task - Investigation Design Section Scientific Writing Assessment

Data Collection and Analysis

All of the assessments designed for measuring science proficiency were administered at the beginning and the end of the year. All assessments were scored using rubrics developed by the research team. 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. The rubric and ICC for each assessment is described below. By using a pre/post administration schedule for these measures, a Paired Samples t Test was employed in analyzing the results for each assessment using the PASW (was SPSS) statistical software package.

Data Sources

<u>*Chemistry Content Knowledge Assessment*</u>: The assessment measured students' ability to know and use scientific explanations of the natural world. The assessment is comprised of eight free response questions, each related to one of several "Big Topics" in Chemistry, as determined by the teachers and researchers. Each question includes an opening paragraph that provides a relevant scenario or context, followed by two questions. One question asks the student to *describe* the fundamental science concept (*Know*) and the other asks the student to *apply* that concept to the scenario provided (*Use*). The rubric for this assessment was developed from answers provided for the questions by an expert chemist with over a decade's 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 team for this assessment achieved an ICC of 0.900.

Scientific Writing Assessment: The scientific writing assessment was developed to assess students' abilities to generate and evaluate scientific arguments. 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 who provides an inaccurate explanation for the data. The students are directed to respond to the scientist's claim by generating an argument in support of a countering claim, which includes evidence and a rationale based on the data and information provided in the question, being mindful of writing style and grammar. The rubric, with an overall possible score of 28 points, was divided into three subscales: *Argument Structure* focusing on the inclusion of fundamental argument components including claims, evidence, and rationale (6); *Argument Content* concerning the quality and relevance of the argument components with respect to scientific discourse (10); and *Mechanics* regarding the punctuation, grammar, and technical quality of the writing (12). The ICC for this scoring team was 0.615.

Chemistry Performance Task Assessment: The performance task assessment was developed to understand and measure the progress in students' abilities to design an investigation that will allow them to generate an argument in response to a research question. The students must develop an original investigation and make decisions about the appropriate data to collect and evidence to use to generate their argument. These assessments are done in groups of 3-4 students, and the group submits a final product for scoring. The final product includes areas for students to describe the investigation they designed, the data they collected, and the argument they created, along with justification for each of these sections. Initial group composition was maintained as much as possible during separate administrations, and if it was not, the resulting scores were not included in the analysis. The rubric for this assessment followed the structure of the assessment packet and focused on technical and theoretical elements present in each section that related to the nature of scientific inquiry. The scoring team for this assessment achieved an ICC of 0.792.

SUSSI: The Student Understanding of Science and Scientific Inquiry (SUSSI) (Liang, Chen, Chen, Kaya, Adams, Macklin, & Ebenezer, 2006) instrument was adapted to measure students' understanding of the development and nature of scientific information. The assessment was comprised of 44 statements about science with Likert-scale agreement responses offered. Analysis of these answers assigned raw points to each response in relation to the nature of the item. Statements representing accurate ideas about science and scientific inquiry were scored a minimum of one point (strongly disagree) to a maximum of five points (strongly agree). Statements representing inaccurate ideas about science were scored in a reverse manner. The authors of this instrument originally separated the assessment into several subscales representing major NOS concepts; however, the researchers grouped these subscales in appropriate groups relating to Aspect 2 of the science proficiency framework. This assessment did not require the use of multiple raters.

Standardized Comparison Content Assessment: 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. This assessment did not require the use of multiple raters.

RESULTS

Table 2 presents the Paired Samples t Test results for the pre/post Chemistry Content Knowledge Assessment, the Scientific Writing Assessment, the Chemistry Performance Task Assessment, SUSSI, and the Standardized Comparison Content Assessment. On the Content Knowledge assessment, the *Know* and *Use* subscales were analyzed. The *Structure, Content,* and *Mechanics* subscales were analyzed for the Scientific Writing Assessment. The Performance Task Assessment was split into subscales concerning *Argument Development* and *Investigation Design*. The SUSSI was split with regard to questions about the *Development* and the *Nature* of science knowledge. Significant differences between Pre and Post scores were found for most of the assessments and the majority of the subscale scores.

Table 2: PASW Output Results for Paired Samples t Tests of Overall and Subscale Assessment Data

Assessment	Pre		Post		*t*	*df*	*p*	*d*
	M	SD	M	SD				
Chemistry Content Knowledge								
Overall	2.34	2.06	9.07	4.92	15.01	105	< .001	1.78
Know	1.46	1.43	5.10	3.05	12.79	105	< .001	1.53
Use	0.88	0.99	3.96	2.42	13.87	105	< .001	1.67
Scientific Writing Assessment								
Overall	15.44	4.65	17.70	4.38	5.062	120	< .001	0.51
Structure	2.64	1.49	3.93	1.44	7.960	120	< .001	0.88
Content	3.79	2.06	4.84	2.11	4.902	120	< .001	0.50
Mechanics	9.01	1.95	8.90	1.89	0.494	120	.622	--
Chemistry Performance Task								
Overall	19.28	2.03	24.74	7.11	5.094	42	p<.001	1.04
Argument Development	4.84	1.29	7.26	1.38	7.153	42	p<.001	1.81
Investigation Design	14.44	2.39	17.49	6.04	3.662	42	.001	0.66
SUSSI								
Overall	87.96	17.21	88.72	13.42	0.375	68	.709	--
Development of Science Knowledge	51.52	12.77	56.33	10.07	3.366	68	.001	0.42
Nature of Science Knowledge	31.72	5.40	32.39	4.71	0.957	68	.342	--
Standardized Comparison Content Assessment								
Overall	6.15	2.15	7.96	2.95	6.396	133	< .001	0.70

Effect sizes for each of the pre/post comparisons were determined by calculating Cohen's *d*, a metric where the categorical range is: small = 0.2, moderate = 0.5, and large = 0.8 or higher (Cohen, 1988). The magnitude associated with these categorizations relates how much of a difference is seen in the mean scores to the normal spread of scores based upon standard deviations. Thus, a significant difference in mean scores that has a large effect size demonstrates that the change that occurred between the pre and post administrations of that assessment represents meaningful development of students' abilities that are the focus of the specific assessment. As the assessments in this study were designed to target aspects of students' science proficiency, statistically significant improvements on an assessment or particular subscale that has an associated large effect size provides evidence for the improvement of students' relevant science proficiencies.

Large effect sizes ($d \geq 0.8$) were seen for all elements of the chemistry content assessment (Total = 1.78; Know subscale = 1.53; Use subscale = 1.67), most of the chemistry performance task (Total = 1.04; Argument Development subscale = 1.81) components, and the Structure subscale on the scientific writing assessment. Moderate effect sizes ($0.5 \leq d \leq 0.8$) were determined for elements of the scientific writing assessment (Total = 0.51; Content subscale = 0.50), the chemistry performance task (Investigation Design subscale = 0.66) and the standardized comparison content assessment (Total = 0.70). A small effect size was seen with the Development of Science Knowledge subscale from the SUSSI assessment ($d = 0.42$) and any comparisons that did not produce significant results were not analyzed for effect size. Figure 3 provides a graphical representation of further effect size analysis through the alignment of assessment components with aspects of science proficiency. This alignment was created by comparing mean changes in scores on different elements of each assessment that related to a specific aspect of science proficiency. For example, elements of the Argument Development subscale scores were used to produce effect sizes for Aspect 2 of Science Proficiency

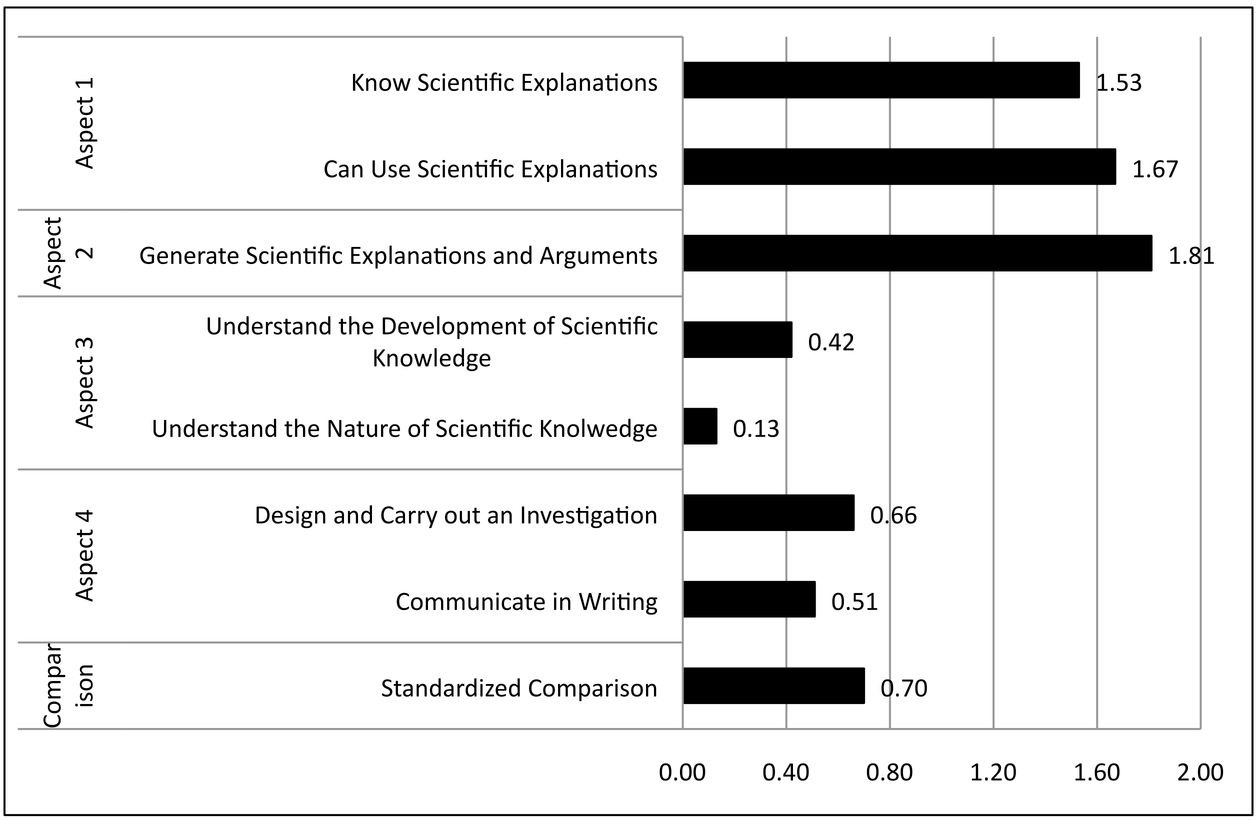

Figure 3: Effect Sizes for Chemistry Assessments Aligned with Aspects of Scientific Proficiency

CONCLUSIONS

The results of this study provide evidence supporting the working hypothesis regarding the positive impact of ADI-based instruction on developing students' science proficiency. The nature of the data and study offers support for a correlational, not causational, relationship between ADI instruction and the development of science proficiency due to the lack of a comparison group. However, national level discussions have established a broad based lack of science proficiency in students (NRC, 2005, 2008), indeed spurring a renewed focus on developing richer and more complex national standards (NRC, 2011) that more appropriately reflect the knowledge and practices of science. The prevalence of such a consensus view among education professionals warrants consideration of significant results although comparison data is not available.

Significant changes from the beginning to end of the year were noted on all assessments administered at the beginning and end of the school year, except the SUSSI instrument. The consideration of the effect size analysis for these different findings provides insight into the magnitude of change, which serves as a reflection of the amount of learning that occurred in the student sample. Over the course of a year, as chemistry students engaged in more authentic and educative investigations facilitated by using the ADI instructional model to design laboratory experiences, they also exhibited meaningful growth in their ability to know and use scientific explanations, generate scientific explanations and arguments, and design and conduct investigations. The findings demonstrate that ADI instruction encouraged the development of some of these elements of high school chemistry students' science proficiency but not others. Understanding of the nature and development of scientific knowledge did not significantly improve for these students over the course of the year. The students' proficiency in communicating through scientific writing did show notable improvement, yet the moderate effect size shows the development of this aspect of science proficiency was more limited than others. Further discussion of specific trends and potential explanations for them are posited below.

Knowing and Using Scientific Explanations

The results for the Chemistry Content Knowledge Assessment display significant and meaningful improvement in students' abilities to know scientific explanations in response to questions asking for their description. Students' ability to also apply that knowledge to explain the events of relevant scenarios and phenomena demonstrated important growth over the course of the year of study. This assessment offers an interesting contrast to the Standardized Comparison assessment results, which also show a significant gain in scores over the year when ADI-based chemistry instruction occurred. The standardized comparison assessment was created using publicly available versions of multiple choice questions from several prominent standardized tests (ex - TIMMS, NY Regents, California Standards). These questions addressed the same content topics that were covered in the Chemistry Content Knowledge assessment and ADI investigations implemented during the school year.

Although the student sample produced a significant improvement in scores, the effect size analysis shows that the gains made on this assessment were only moderate. Comparing this to the large effect size improvement observed in the more challenging assessment, the notion that typical standardized assessments are accurate measures of student abilities could be challenged. The rubric for the Content Knowledge assessment was developed using the answers of expert chemists, thus the students' responses were judged and scored against that standard, whereas with the standardized assessments, the responses are scored for alignment with canonical knowledge of chemistry. The improvement in the mean score for the *Use* subscale was three points, which is a notable improvement in a mean drawn from a sample of over a hundred students, particularly in light of the written nature of these responses. Standardized assessments rarely offer an opportunity to accurately measure this facet of this science proficiency aspect. As change efforts and standards movements increasingly focus on the development of a broader array of abilities in students, the assessments that will drive many administrative and instructional decisions must concordantly be improved to more accurately target those enhanced learning goals. Assessments in the vein of the Chemistry Content Knowledge assessment provide examples of what those improved assessments could look like.

Communicating through Scientific Writing

The results for the scientific writing assessment support that engagement in more authentic and educative practices, including the creation of persuasive lab reports and participation in peer review of others' reports coincides with the overall improvement of students' science writing proficiency. However, the magnitude of this improvement was not as great as the improvements observed in other aspects of science proficiency from this group of students. The limited nature of this specific result can be attributed in part to contextual factors which arose in the chemistry classrooms that are the focus of this study. Students in this course are mostly members of the 11^{th} grade student body typically found in US schools. During the year of this study, this grade level is also the grade where the high school level, high-stakes standardized test is given, whose scores are used in calculating AYP scores and other benchmarks of school system success.

During the second half of the year, the teachers who taught this course had to limit the amount of ADI laboratory experiences their students engaged in so as to allow for more time to conduct comprehensive reviews of biology, chemistry, and physics content material that could be addressed on the standardized test. These teachers also chose not to have students create lab reports for any ADI activities they did engage in during that time to lessen the pressure of assignments as students' prepared for testing. Thus, the limited amount of writing opportunities for these students could have contributed to the lessened magnitude of development of this aspect of science proficiency. Indeed, data not shown here but collected from a unique administration of this assessment during the middle of the school year demonstrates that the majority of improvements in writing scores observed for this year occurred during the first semester of the year, when students engaged in fully implemented ADI investigations, including writing lab reports and peer review sessions.

Understanding the Nature and Development of Scientific Knowledge

The one aspect of science proficiency that did not appear to develop significantly over the year of this study concerns students' ability to understand the nature of scientific knowledge and how that knowledge is developed. The SUSSI assessment (Liang et al., 2006) used to measure this aspect of science proficiency was developed through an extensive validation effort involving US, Chinese, and Turkish classrooms. The concepts and ideas that comprise the content of this assessment are drawn from a well-established body of research more commonly recognized as Nature of Science/Nature of Scientific Inquiry (NOS/NOSI) education (Hodson, 2009; Lederman, 1992, 2004; McComas, 1996). The limited nature of the development observed from this student sample is not surprising in light of findings from other research that has established these more abstract ideas as particularly challenging for students to learn and understand (Hodson, 2009; NAS, 2008).

As this study is part of an exploratory study aimed at refining the ADI instructional model through an evidence-based iterative process, the results concerning this aspect of science proficiency did provide evidence for the research team in considering changes to the model. NOS/NOSI research literature has established the need for explicit and reflective discussion regarding the concepts that would fall into this categorization (Abd-El-Khalick, Bell, & Lederman, 1998; Akerson, Abd-El-Khalick, & Lederman, 2000). To accommodate this requirement in the ADI instructional model, Stage 7 (refer to Figure 1) involves an explicit and reflective discussion that incorporates not only the science content addressed during a particular investigation, but also should address the relevant NOS/NOSI topics for an investigation as well. In light of the disappointing findings from the SUSSI assessment and deliberations with the participating teachers in the project, the research team determined that intentionally placing this stage of activity towards the end of the model was too late for meaningful learning and discussion to occur. In response, this stage of activity has been purposefully moved to Stage 5, directly following the argumentation session and preceding the initial drafting of the investigation report. The project teachers also felt this was a more sensible placement since they observed that these kinds of discussions more naturally emerged at this point in the model. The second year of data collection will provide insight as to the impact of this change.

IMPLICATIONS

The study described in this proposal provides further evidence of the benefits of argument focused science instruction. The findings further the research base on the impact of specific argument focused curricula and potential targets for improvement. The use of argumentation and instructional models that privilege these kinds of interactions can be useful in teaching not only science content, but also enhancing students' abilities in other areas important to learning and understanding science. By engaging in authentic science practices, such as designing and conducting investigation, defending and critiquing scientific arguments, and coordinating claims, evidence, and justifications into scientific arguments communicated through writing, students will not only enhance their potential to be scientists, but also improve their status as a global citizen with proficiency in understanding and evaluating that vast amount of information and public opinion increasingly shaped by science.

As K12 science education shifts focus to developing students' science proficiency, this study contributes to the research base on ways of assessing aspects of a very broad and complex construct that serves as one approach to understanding the learning of critical thinking skills. The broad performance profile generated from this analysis also demonstrates the importance of using multiple assessments for gaining insight into complex science learning, as opposed to more direct and common standardized measures. In light of the conference theme, *Non Satis Scire: To Know Is Not Enough*, the need and drvie towards science proficiency, including a broad array of abilities beyond simple knowledge of science content, exemplifies this ideal. As such, the assessments that will always shape the development of educational systems must come to reflect those same ideals and be able to provide a collection of measures that truly captures the aspects of science proficiency.

Acknowledgements

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

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education*, 82(4), 417-436.

Akerson, V. L., Abd-El-Khalick, F., & Lederman, N. G. (2000). The influence of a reflective activity-based approach on elementary teachers' conceptions of the nature of science. *Journal of Research in Science Teaching*, 37(4), 295-317.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*, 141-178.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. Science Education, 84(3), 287-313.

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

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.

Hodson. (2009). *Teaching and learning about science.* Rotterdam, The Netherlands: Sense Publishers.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. Science Education, 77(3), 319-337.

Lederman, N. (1998). The state of science education: Subject matter without context. *Electronic Journal of Science Education*, ISSN 1087-3430, 3(2). Retrieved from http://unr.edu/homepage/jcannon/ejse/ejsev3n2.html

Lederman, N. G. (2004). Syntax of nature of science within inquiry and science instruction. In L. Flick and N. Lederman (Eds.), *Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education.* (p. 301 – 317). Netherlands; Kluwer Academic.

Liang, L., Chen, X., Chen, S., Kaya, O., Adams, A., Macklin, M. & Ebenezer, J. (2006). Student Understanding of Science and Scientific Inquiry (SUSSI): Revision and Further Validation of an Assessment Instrument. *Paper presented at the 2006 Annual Conference of the National Association for Research in Science Teaching (NARST)San Francisco, CA, April 3-6.*

McComas, W. (1996). Ten myths of science: Reexamining what we think we know. *School Science & Mathematics, 96*, 10.

National Academy of Science. (2008). *Science, evolution and creationism.* Washington, DC: National Academy Press.

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.

National Research Council. (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, D.C.: National Academies Press.

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A.J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263-305). Mahwah, NJ: Erlbaum.

Roberts, D.. (2007). Scientific Literacy/Science Literacy. In S. Abell & N. Lederman (Eds.), *Handbook of research in science learning* (p. 729-780). New Jersey: Lawrence Erlbaum.

Sampson, V., & Clark, D. (2011). A comparison of the collaborative scientific argumentation practices of two hIgh and two low performing groups. *Research in Science Education 41(1),* 63-97.

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.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*(3), 345-372.

The Design-Based Research Collective. (2003). Design-Based Research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.