NEW INSTRUMENTS THAT CAN BE USED BY RESEARCHERS TO ASSESS THREE DIFFERENT ASPECTS OF SCIENCE PROFICIENCY

Victor Sampson, Jonathon Grooms, Patrick Enderle

Background/Context
Scientific proficiency, as described by (Duschl, Schweingruber, & Shouse, 2007), encompasses 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.

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 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. However, the dilemma of assessing students’ scientific proficiency or measuring changes in science proficiency over time still prevails. Typical paper/pencil, multiple choice type assessments are not adequate for assessing the many facets of scientific proficiency (Duschl et al., 2007; National Research Council, 2001, 2008). New assessments, therefore, are needed in order assess students’ current levels of scientific proficiency and how they develop different aspects of science proficiency over time.

Purpose/Objective/Research Question/Focus of Study
This proposal describes three new and innovative assessments that are being used within a larger study to measure students’ scientific proficiency. 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. Using a single assessment or inappropriate assessments over estimate or even under estimate students’ scientific proficiency depending on the nature of the assessment. The three assessments, which will be discussed in this proposal, were designed to measure students’: (1) understanding of science content and their ability to use those understandings to explain; (2) understanding of the nature and development of scientific knowledge; and, (3) abilities to generate and evaluate scientific arguments.

Setting and Participants
The assessments described in this proposal are being pilot-tested by six teachers ranging from grade seven through grade twelve within a K-12 University Research School. This pilot testing is taking place 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 pilot testing of the model and the assessments are taking place in middle
school life science, middle school physical science, high school biology, and high school chemistry courses. There are approximately 600 student participants involved in this research project, whose responses will be used to determine the validity and reliability of the assessments.

**Intervention/Program/Practice**

The broader context of this research is aimed at refining the Argument-Driven Inquiry instructional model so that teachers can use it within the context of an existing middle or high school science curriculum to provide a high quality laboratory experience for their students. 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. As part of this project, the three assessments of science proficiency discussed in this proposal are being administered on three occasions during the school year, once at the beginning of the year, again at the mid-point of the school year, and finally at the conclusion of the school year. This pre-, mid-, post-assessment strategy allows the researchers to track students’ progress over the course of the school year and to measure how their levels of different aspects of science proficiency change over time. The assessments used as part of this study, and their development, will be described in the following section.

**Research Design**

The assessments discussed in the proposal focus on three of the four key aspects of scientific proficiency. Students’ abilities to participate in the practices and discourse of the scientific community in a productive manner, the fourth aspect of science proficiency, is inherent in each of the assessments; however, it is not easily assessed using a paper/pencil style assessment. This aspect of scientific proficiency is being assessed using the Assessment of Scientific Argumentation in the Classroom observation protocol, which has already been developed and validated (Enderle, Walker, Dorgan, & Sampson, 2010). Table 1 identifies the aspects of science proficiency and the accompanying assessment that will be addressed in this proposal. The three assessments were developed by the research team to both align with particular aspects of scientific proficiency as well as the “Big Ideas” of a particular domain of science. Therefore, four versions of each assessment were created to accommodate the Big Ideas of each subject area that is a part of this study: middle school life science, middle school physical science, high school biology, and high school chemistry. Furthermore, to account for potential testing effects, three slightly different versions of each assessment were created to use during the pre-, mid-, and post-intervention data collection periods. This proposal addresses the general features of each type of assessment and uses high school biology as a representative sample of all the assessments.

<table>
<thead>
<tr>
<th>Aspect of Science Proficiency</th>
<th>Assessment Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students know, use, and can interpret scientific explanations of the natural world</td>
<td>Topic-specific Content assessment</td>
</tr>
<tr>
<td>Students can generate and evaluate scientific explanations and arguments</td>
<td>Topic-specific Writing assessment</td>
</tr>
<tr>
<td>Students understand the nature and development of scientific knowledge</td>
<td>Topic-specific Performance Task</td>
</tr>
</tbody>
</table>

**Topic-specific Content Assessment.** Each content knowledge assessment is comprised of eight free response questions, with a question related to each Big Idea identified for the corresponding course. Each question begins with the presentation of a background scenario and/or information coupled with some type of graphical representation (ex – a graph of reaction data, an image of the organism discussed, a map of an area described, a diagram of a system, etc.). This information provides a specific context for the students to apply their understanding of the related Big Idea.
Following the opening paragraph in each question, students are presented with a set of two questions. The first question in each set asks the student to describe the science concept underlying the question related to the Big Idea, followed by the second question, which asks the student to apply their understanding of that science concept in order to explain the information provided in the introductory scenario. Structuring the questions in this way allows us to assess the know and use aspects of scientific explanations, which as noted earlier, are an indicator of science proficiency. Figure 1A provides a sample question from the high school biology content assessment.

Once the content knowledge assessments were fully developed, an expert consultant reviewed the assessments suitable to their expertise. The consultants were asked to verify the appropriateness of the questions in relation to the associated Big Idea; and they were also asked to provide an “expert answer” to each of the questions. These answers were used to construct itemized rubrics that will be used to score student responses to each question.

Figure 1. (A) A sample item from the high school biology content assessment, (B) a sample prompt from the high school biology scientific writing assessment, and (C) a sample task from the high school biology performance task assessment

**Topic-specific 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/doctor who provides an inaccurate explanation for the data. The students are then directed to respond to the scientist’s claim by arguing in support of a countering claim, being sure to include evidence and a rationale as part of their argument based on the data and information provided in the question, all-the-while being mindful of writing style and grammar.

During this assessment the students are provided with several pieces of lined paper to help organize their writing. The students are initially asked to engage in a pre-writing activity to outline their argument and then generate a rough draft. By the conclusion of the assessment period the students are expected to have refined any initial drafts or pre-writing exercises in order to provide a final draft of their argumentative essay addressing the task identified for the assessment. A sample prompt from a topic-specific writing assessment is provided in Figure 1B.

**Topic-specific Performance Task.** The performance task assessment was developed to understand and measure the progress in students’ abilities to develop an investigation in response to a research question, making decisions about the appropriate data to collect and evidence to use in creating an argument that answers the research question. By completing a task like this students are able to demonstrate, in an authentic context, their understanding of the nature and development of scientific knowledge, as well as
their ability to generate scientific arguments and explanations. For each course, the performance tasks were developed such that they were not specifically related to any of the Big Ideas or curricular standards, but dealt with material from the discipline, thus providing a novel context for the assessment. These assessments are done in groups of 3-4 students, and the group submits a final product for scoring. Groups are provided with a short paragraph describing some of the science concepts underlying the activity (ex – harmonic motion, comparative anatomy). Following this, a research question is given along with a list of materials that are available at the physical station where the group is performing this assessment. The assessment packet then has a set of open-ended questions, which ask the students to record and explain their decision-making processes and the results of them. Once students have completed the task and developed their groups’ argument, they are asked to highlight particular elements of their argument for scoring clarity. An example of a topic-specific performance task is provided in Figure 1C.

Data Collection and Analysis
At the time of this proposal the data collection process is currently in progress, with the post-intervention data to be collected in late May 2011. Once the data set is complete the research team will use the rubrics that have been developed to assess the validity and reliability of the three assessments. During that process inter-rater reliabilities and item analyses will be determined in addition to the construct validity that has been generated by the content expert reviewers. It is anticipated that the complete data set will represent approximately 600 students, each having completed the three assessments above. The students’ scores on these assessments will provide insight as to their levels of scientific proficiency with regard to the specific characteristics discussed above as well as providing measures of how their scientific proficiency has changed as a result of the ADI intervention.

Findings/Results
We will be able to report on the validity and reliability of the new instruments once data collection and analysis is complete in June 2011.

Conclusions/Potential Implications
We anticipate that the three new assessments will provide valuable data to help measure students’ scientific proficiency, which is an important outcome of science instruction. The use of assessments such as these in other research project will provide a better measure of what students know and are able to do than traditional multiple-choice tests. Given the anticipated findings based on these assessments, we hope to show that similar assessments will be useful in measuring complex constructs that are valued in large-scale research projects.

References