CrossMark

Assessing epistemological beliefs of experts and novices via practices in authentic science inquiry

Melanie E. Peffer[1*] and Niloofar Ramezani[2]

Abstract

Background: Achieving science literacy requires learning disciplinary knowledge, science practices, and development of sophisticated epistemological beliefs about the nature of science and science knowledge. Although sophisticated epistemological beliefs about science are important for attaining science literacy, students' beliefs are difficult to assess. Previous work suggested that students' epistemological beliefs about science are best assessed in the context of engagement in science practices, such as argumentation or inquiry.

Results: In this paper, we propose a novel method for examining students' epistemological beliefs about science situated in authentic science inquiry or their Epistemology in Authentic Science Inquiry (EASI). As a first step towards developing this assessment, we performed a novice/expert study to characterize practices within a simulated authentic science inquiry experience provided by Science Classroom Inquiry (SCI) simulations. Our analyses indicated that experts and novices, as defined by their experience with authentic science practices, had distinct practices in SCI simulations. For example, experts, as compared to novices, spent much of their investigations seeking outside information, which is consistent with novice/expert studies in engineering. We also observed that novice practices existed on a continuum, with some appearing more-or less expert-like. Furthermore, pre-test performance on established metrics of nature of science was predictive of practices within the simulation.

Conclusions: Since performance on pre-test metrics of nature of science was predictive of practices, and since there were distinct expert or novice-like practices, it may be possible to use practices in simulated authentic science inquiry as a proxy for student's epistemological beliefs. Given than novices existed on a continuum, this could facilitate the development of targeted science curriculum tailored to the needs of a particular group of students. This study indicates how educational technologies, such as simulated authentic science inquiry, can be harnessed to examine difficult to assess, but important, constructs such as epistemology.

Keywords: Epistemology, Nature of science, Authentic science inquiry, Educational technology

Introduction

Science literacy is multidimensional and includes general knowledge, understanding of practices such as argumentation or inquiry, positive attitudes towards and experiences in science, development of appropriate mental models about complex relationships, and epistemological beliefs about the nature of science and generation of science knowledge (Renken et al. 2016; Elby et al. 2016; Schwan et al. 2014).To properly educate a scientifically literate populace, it is necessary for science education to include all of these facets; however, many facets overlap, and some, such as authentic inquiry and epistemological beliefs about science, are pedagogically challenging. For example, although inquiry is an essential science practice for generating new science knowledge students are overwhelming exposed to simple, rather than authentic, science inquiry in K-16 classrooms (Chinn and Malhotra 2002). Furthermore, some have suggested that each discipline of science has its own distinct nature of science (NOS) principles (Schizas et al. 2016). If there are different ways of conceptualizing science inquiry, a key question for science educators is as follows: to attain a

* Correspondence: melanie.peffer@unco.edu
[1]College of Natural and Health Sciences, University of Northern Colorado, Ross Hall 1556, Campus Box 92, Greeley, CO 80639, USA
Full list of author information is available at the end of the article

scientifically literate populace, which type of inquiry, NOS principles, and epistemological beliefs about science do we teach in a classroom? If consensus regarding the nature of inquiry is difficult to attain, how can science educators effectively evaluate and assess epistemological understanding? This paper proposes a novel method for assessing epistemology situated in authentic science inquiry. As a first step towards developing a formal assessment, we attempt to define practices that align with what experts do in authentic science inquiry and established metrics of nature of science (NOS) understanding and epistemological beliefs about science.

Overlap and distinction between NOS, epistemology, and authentic science inquiry

Nature of science

Lederman et al. (2002) stated that "NOS refers to the epistemology and sociology of science, science as a way of knowing, or the values and beliefs inherent to scientific knowledge and its development" (p. 498). The authors go on to say that important aspects of NOS to focus on pedagogically include the tentative nature of science knowledge, that science knowledge is empirically generated and influenced by the social and cultural environment from which it was generated, and involves human inference and creativity. The authors operationalized NOS as the epistemological aspects that underlie scientific processes such as inquiry, but are not the same as the processes themselves. In a later paper from the same research group, the nature of science inquiry is redefined to include that scientific inquiry begins with a question and follows a non-linear path that can vary extensively and that inquiry may or may not include hypothesis testing. Moreover, data are not equivalent to evidence and explanations for a phenomenon reconcile the data collected by the investigator with what is already known (Lederman et al. 2014).

Although Lederman's definition of NOS is commonly used throughout the literature, it is subject to debate and critique. For example, an individual's understanding and perception of NOS may change over time, either in response to changes in the field or personal experiences (Deng et al. 2011). There is also a question of whether or not there is a universal, domain-general NOS (Abd-El-Khalick 2012) or if NOS is better conceptualized within the context of a domain (Schizas et al. 2016). In addition to potential disciplinary differences, there is a wide variety of interpretations of NOS among practicing scientists, both within and between various disciplines (Sandoval and Redman 2015; Schwartz and Lederman 2008).

Epistemology

Epistemology is the study of what knowledge is and the exploration of what it means to know something.

Questions of epistemology include exploring the nature of truth, justification, and how knowledge manifests as skills versus facts (Knight et al. 2014). In the context of science education, most prior work has focused on NOS understanding and personal epistemology (Elby et al. 2016). Deng et al. (2011) contended that epistemology cannot be separated from NOS as is an essential component of inquiry practice. Alternatively, NOS as a term could also be considered interchangeable with personal epistemology or epistemic cognition (Greene et al. 2016). Personal epistemology is the set of views or beliefs (known as epistemic beliefs) a person has about the nature of knowledge and knowing (Elby et al. 2016; Schraw 2013). Personal epistemology is thought to contain cognitive structures, one of which is epistemological commitments. Zeineddin and Abd-El-Khalick (2010) suggest that epistemological commitments can influence how students reason about a science problem and may explain the disconnection between how one thinks in a formal science context versus their day-day life.

Hofer and Pintrich (1997) characterized four dimensions of scientific epistemic beliefs: certainty, development, source, and justification. Source and justification both deal with the nature of knowledge and knowing. In the case of science, a less sophisticated belief about the source of science knowledge is that it is derived from an outside authority person, rather than by resulting from one's inquiry strategies (Conley et al. 2004). Justification in science is how an individual uses data or evidence, particularly generated through experiments, to support their claims. Certainty refers to the nature of knowledge as concrete versus tenuous. Certainty is also found in NOS theory; for example, someone with a less sophisticated understanding of certainty (and/or poor understanding of NOS) would claim that scientific knowledge is certain and unchanging and that it is possible to obtain a single "correct" answer. This belief also relates to the development domain in that a more sophisticated understanding would be that scientific information could change in light of new developments (Conley et al. 2004). In the context of authentic science inquiry environment provided by Science Classroom Inquiry (SCI) simulations (Peffer et al. 2015), we can observe and analyze how a student engages with source (where do participants look for information, and why?), justification (how data are used to support claims?), certainty (how do the students discuss their results?; already examined by Peffer and Kyle 2017), and development (how does the student's interpretation of the problem change in light of new information?).

Authentic science inquiry

The Next Generation Science Standards (NGSS; NGSS Lead States 2013) replaced the teaching of inquiry as a

standalone phenomenon with scientific practices. Practices are defined as "a set of regularities of behaviors and social interactions that, although it cannot be accounted for by any set of rules, can be accounted for by an accepted stabilized coherence of reasoning and activities that make sense in light of each other" (Ford 2015, p. 1045). Practices include competencies such as engaging in evidence-based argument, modeling, and inquiry. The pedagogical emphasis shifts away from presenting inquiry as a set of rules, such as a prescribed scientific method, and instead focuses on understanding inquiry in its social and cultural context as well as its relationship to other practices. Engaging students in the contextual practices of science is thought to lead to improved science literacy by promoting the development of an understanding of the epistemic process of science, namely generation and nature of science knowledge (Osborne 2014b). Since these practices are intertwined, some specifically study the intersection between practices, such as model-based inquiry. Model-based inquiry is defined as "an instructional strategy whereby learners are engaged in inquiry in an effort to explore phenomena and construct and reconstruct models in life of the results of scientific investigations" (Campbell et al. 2012, p. 2394). Windschitl et al. (2008) suggested that model-based inquiry provides a more epistemologically authentic view of inquiry as it involves five epistemic features related to science knowledge (testable, revisable, explanatory, conjectural and generative) that are often missed with a focus on the type of inquiry promoted by the scientific method. Whether or not an underlying model is required for inquiry is best determined within the context of the inquiry experience and overall pedagogical or research goals.

Inquiry is the predominant means used by scientists to generate science knowledge; however, there is no single definition of inquiry. Hanauer et al. (2009) proposed that scientific inquiry exists on a continuum and that inquiry is best operationalized through the context and overall pedagogical goals. For example, the pedagogical goal of the inquiry experience could be to develop new knowledge or the goal could be for the student to gain personal and cultural knowledge about the process of inquiry. Similarly, Chinn and Malhotra (2002) claimed that science inquiry exists on a continuum based on scientific authenticity. They define authentic inquiry as the complex activity that scientists perform to generate new scientific knowledge. Chinn and Malhotra (2002) listed cognitive processes involved in inquiry such as the generation or research questions, engaging in planning procedures, or finding flaws within experimental strategies and compare these processes between authentic and simple inquiry. For example, in the case of generating research questions, in authentic inquiry scientists generate their own questions, whereas in simple inquiry a research question is provided to the student.

A meta-analysis of research on authentic science inquiry determined that the most common way of defining authenticity in science is providing students the experience "of what scientists 'do' (practices), how science is done, and what science 'is'" (Rowland et al. 2016, p. 5). However, authentic inquiry does not require hands-on experiences. For example, if a hands-on lab experience is used to replicate a known phenomenon to a student, it may be hands-on, but is clearly what Chinn and Malhotra (2002) refer to as simple inquiry since a research question is provided to the study, the student follows simple directions, and there is no planning involved. The question of hands-on experiences and authenticity is also reflected in differences in inquiry practices both between domains of science as well as between practicing scientists within those disciplines (Schwartz and Lederman 2008; Sandoval and Redman 2015). For example, in the context of biology, an individual who uses bioinformatics may study gene expression, but would not work in a "wet" laboratory space but instead solely on a computer. This is in contrast to a biologist who works in a "wet" laboratory studying gene expression who engages in hands-on activities on a daily basis. As an example of the variety of practices observed between domains, a chemist may be engaged in stereotypical bench research with reagents and flasks, but an astronomer may be engaged in observational studies. All are examples of authentic science, but the authenticity is not derived from exactly what each is doing on a day-to-day basis. Instead, authenticity is derived from cognitive processes such as generation of new research questions, use of complicated procedures, and a lack of a single correct answer. In the context of NOS theory, Abd-El-Khalick (2012) suggested that domain-general versus specific questions are best addressed within specific areas of research, such as authentic inquiry practices. Like NOS, characterization of authentic inquiry should reflect both accepted disciplinary practices and the overall pedagogical goals in which the experience is situated.

Examples of authentic inquiry experiences for students could include exposure to course-based undergraduate research experiences or CUREs (Auchincloss et al. 2014; Corwin et al. 2015) or simulated authentic science inquiry such as the SCI simulations (Peffer et al. 2015). Although "simulated" and "authentic" may seem contrary to one another, we would argue that in the case of SCI simulations the simulated nature of the experience does not detract from its authentic features because it models the thought process used by scientists when engaged in an unstructured, real world problem. Simulated experiences are beneficial when considering how computer based experiences can be leveraged for high-throughput assessment. Furthermore, as discussed above, 'hands-on' is not

synonymous with 'authentic' when categorizing inquiry experiences. There is no single definition for a simulation, and in fact simulations exist on a continuum from rigid algorithmic models of some aspect of reality to modeling a real phenomenon (Renken et al. 2016). The simulated authentic inquiry experience provided by SCI is best characterized as a conceptual simulation (Renken et al. 2016) because it models an abstract process, namely authentic science inquiry, in a scaffolded, autonomous manner using real-world problems and data. SCI is an authentic experience because it models the thought process and decision-making used by scientists engaged in unstructured real world problems. Although the inquiry experience exists on a computer, it still maintains many facets of authentic inquiry as defined above, allowing students to generate new evidence based ideas and knowledge and requiring students to engage in a non-linear process that is not directed at a single correct answer.

Connecting NOS and epistemology to authentic science inquiry

In the context of authentic science inquiry, we argue that understanding of NOS or NOS-inquiry and epistemological

beliefs about science are intertwined and influence both each other and science practices (Fig. 1). As pointed out by Elby et al. (2016), divisions between NOS and personal epistemology research may stem from the separate nature of the two literatures, as both are related to understanding how individuals conceptualize the nature of science knowledge, yet are published in journals targeting different readership, namely science educators versus psychologists. What a student knows about science and authentic science inquiry will influence what they believe about science. For example, the tentative nature of science knowledge is a NOS principle identified by Lederman et al. (2002) and certainty of knowledge is one of the dimensions of science epistemology identified by Conley et al. (2004). If a student understands that science knowledge can change in light of new evidence, they likely also believe that science knowledge is not certain. Conversely, if a student has an epistemological belief that knowledge changes over time, then it would be easier to learn the principle that science knowledge is tentative. Furthermore, encouraging students to memorize tenants of NOS understanding out of context does not necessarily enhance a students' science literacy (Deng et al. 2011). Said otherwise, a student can memorize
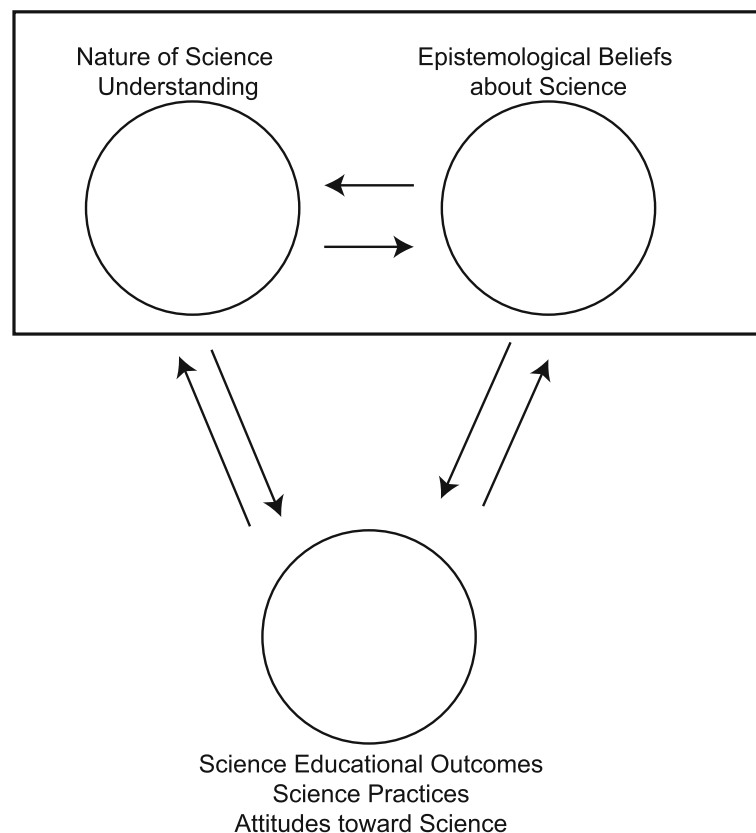

Fig. 1 Our conceptualization of the relationship between NOS understanding, epistemological beliefs, and outcomes. We propose that what individuals know and believe influence one another, and that these can then influence practices. An individual's experience in a classroom or with a practice can also influence what they know and believe as well

NOS tenants, but these tenants may not translate into sophisticated epistemological beliefs or practices, or transferable skills that will be useful to students in the real world. How epistemological beliefs about science relate to inquiry practices are not well understood (Sandoval 2005). This leads to a central question of this work: how is understanding of NOS and epistemological beliefs enacted as student practices in authentic inquiry? For example, our previous work suggested that experts using more tentative language when making their conclusions (Peffer and Kyle 2017). This may indicate that experts use more tentative language because they understand that scientific knowledge is not concrete and subject to revision in light of new evidence. Therefore, analysis of inquiry practices, particularly practices in authentic inquiry, may provide a new assessment strategy of NOS/epistemology. We now turn to a discussion of current assessments of NOS/epistemology and the potential of using practices instead of conventional assessments.

Current assessments of NOS/epistemology

NOS understanding and sophisticated science epistemic beliefs are an essential part of both science education and science literacy (Renken et al. 2016) and many assessments of NOS or epistemology have been developed (Akerson et al. 2010; Conley et al. 2004; Koerber et al. 2015; Lederman et al. 2014; Lederman et al. 2002; Stathopoulou and Vosniadou 2007). However, definitions of both NOS and epistemological beliefs about science are difficult to concretize and operationalize leading to assessment challenges. For example, which or whose NOS understanding do we want students to adopt? Which aspects are most crucial for science literacy? What defines a "sophisticated" epistemological belief about science and how does this vary both within and outside of various science disciplines? By trying to fit participants neatly into categories such as "sophisticated," we lose the breadth of information associated with the wide variety of ways that NOS can be operationalized (Sandoval and Redman 2015). Another limitation is that current assessments of NOS and epistemology are taken at one fixed point in time and do not reflect changes in understanding of these principles over time. Forced choice assessments assume that we can neatly fit participants into categories that match the philosophical beliefs of the survey authors, giving a limited view of how the student conceptualized their understanding of NOS or their epistemological beliefs about science (Sandoval 2005; Sandoval and Redman 2015). Furthermore, Likert scale metrics, often used in these assessments, are criticized for their lack of reliability and validity, and some have called for a cessation of the use of these metrics (Sandoval et al. 2016).

Examining practices as assessment of epistemology

One possible solution to assessment challenges is to examine student science practices in real time. For example, Deng et al. (2011) suggested that sophisticated NOS understanding should be interpreted in the context of how well students argue scientific claims. Sandoval (2005) observed that how specific epistemological beliefs relate to inquiry practices is largely unknown and suggested that to understand how students make sense of science, an essential research focus was to examine their practices in authentic science inquiry. In the context of evidence evaluation, part of both science inquiry and important for overall literacy, Chinn et al. (2014) proposed the Aims and Values, Epistemic Ideals, Reliable Processes (AIR) model for examining epistemic cognition in practice. AIR stands for Aims and Value, Epistemic Ideals, and Reliable Processes and was designed to reflect three different aspects of epistemic cognition. Epistemology has also been examined in practice in the context of how sixth grade students evaluate and integrate online resources (Barzilai and Zohar 2012).

Expert/novice studies to define assessment criteria in SCI

One way to connect practices to a sophisticated NOS/epistemology is to examine what experts and novices do in authentic situations. Defining what experts are doing (as compared to novices) in the context of authentic science inquiry could lead to the development of criteria that could be used for detecting differences among novices (i.e., who has more-or-less expert-like practices in a group, and by proxy differences in epistemology) and potential areas for pedagogical intervention. Expert/novice differences in practices have been examined in the context of engineering education. Atman et al. (2007) examined professional engineers working in their field and undergraduates majoring in various engineering sub-disciplines. Participants were tasked with designing a hypothetical playground in a laboratory setting while verbally describing their process. The authors observed that experts engaged in the task for longer, particularly when researching the problem at hand, and gathered both more and a greater variety of information during their activity. Worsley and Blikstein (2014) compared students with either bachelors or graduate degrees in engineering against students without formal training in engineering to design a stable tower that could hold a 1 kg weight. They observed that expert students tended to engage in iterative strategies, repeatedly testing and refining their designs, and returning to planning throughout the process, rather than one stage of planning at the beginning observed in novice students. Although expertise was defined differently in each study, both studies observed a similar trend that experts tended to have similar iterative processes that involve a mix of doing

and refining/seeking additional information. This result may parallel practices in authentic science inquiry, where it is necessary to preform both investigative and information seeking actions as part of a larger investigation. The variability in practices between experts and novices suggests there is an underlying explanation, such as different epistemological beliefs about the practice of engineering design that may explain these behaviors. Within the domain of physics, Hu and Rebello (2014) found that students' epistemological framing, particularly around the use of math in solving physics problems, influenced whether they approach physics problems in a more expert-like manner. The authors found that students presented with hypothetical debate problems instead of conventional physics problems were more expert like in their solutions, focusing on qualitative and quantitative sensemaking instead of on how to plug numbers into a memorized equation (Hu and Rebello 2014).

Current study

If understanding the connection between epistemological beliefs as they relate to inquiry is key to understanding how students make sense of science (Sandoval 2005), it is important to examine student practices in authentic science inquiry. A decade after first calling for examining student practices as a proxy for epistemological beliefs, Sandoval and colleagues reiterated this point and called for further research on how epistemological beliefs can influence certain behaviors, such as how epistemological beliefs underlie the interpretation and analysis of scientific information found on the internet. Since NOS and epistemological beliefs about science are intertwined in inquiry (Fig. 1), and at least in engineering, expert and novice practices varied along predictable lines, we suggest that expert and novice practices in authentic science inquiry may also be indicative of underlying NOS and epistemological beliefs. Furthermore, defining what experts do in simulated authentic inquiry as compared to novices may lead to some consensus of what constitutes sophistication. Using practices in authentic science inquiry as a proxy for students' underlying NOS understanding and epistemological beliefs about science, the researcher or instructor can view what students do in an autonomous, open-ended environment, rather than retroactively assessing students at a single time point.

To develop a practices-based assessment of epistemology/NOS understanding, which identifies expert criteria as the benchmark for future assessment, we examined the practices of experts and novices in the simulated authentic science inquiry environment provided by SCI simulations and connected inquiry practices to existing metrics of NOS and epistemological beliefs about science. Since we

argue that NOS understanding in the case of inquiry cannot be separated from epistemology, we refer to the participant's putative epistemology/NOS understanding as seen through their inquiry practices as their *Epistemology in Authentic Science Inquiry* or EASI. Given the limitations of existing metrics of NOS and science epistemology, this study lays a foundation for the novel use of simulations to assess constructs such as NOS and epistemology. Using student practices embedded in an authentic activity as a proxy for their underlying beliefs mitigates concerns about constraining what students know and believe to a static metric, which is a limitation of existing pen and paper surveys (Sandoval and Redman 2015). The SCI simulation engine also permits autonomy for participants to engage with the simulation in a variety of ways demonstrating the wide range of practices that exist between novices and experts. By assessing EASI via practices, the diversity of approaches utilized by students could lead to the development of systems or curricula that are personalized to individual students. To examine differences in EASI between experts and novices, and how practices in authentic inquiry relate to previously established metrics of NOS/epistemology, we utilized a mixed-methods approach (Creswell 2014), to address the following research questions:

1. What distinguishes expert (meaning, a trained biologist) from novices (undergraduate students) on established metrics of NOS and epistemology?
2. What distinguishes the authentic science inquiry practices of an expert versus a novice?
3. How does expert/novice performance on established metrics relate to authentic science inquiry practices?

Methods

Participants

There were 28 total participants in this study: 20 novices and 8 experts. All participants were associated with a university in a large southeastern city in the USA. The 20 novices were all undergraduate students, mainly in their third or fourth year of college (70% seniors, 20% juniors, 10% sophomores, no freshman) with little, if any, experience in authentic science practices. A single novice participant stated that they had been working in a psychology lab for the last 2 years and had presented a research poster showcasing their work. However, this participant was not listed as an author on any primary research manuscript. For reproducibility purposes, we have marked this novice in Table 2 with a double asterisk. No other novices indicated any *bona fide* experience with authentic science practices. Expertise was defined based on experience with authentic science practices, particularly in biology or related fields (neuroscience,

public health), although none were experts in ecology or conservation biology, which was the topic of the SCI simulation completed. All experts had engaged in biological sciences research for at least 2 years and were listed as authors on primary research manuscripts either submitted or published at the time of the study. The majority of experts were advanced doctoral students, defined by their completion of their comprehensive examinations and achieving candidacy. A single expert had an earned PhD (marked in Table 2 with an asterisk) and was currently working as a postdoctoral associate. Novice participants had diverse ethnic backgrounds. Fifteen percent were white/European-American, 45% black/African-American, 30% Asian-American, 5% multi-racial, and 5% other. The expert population was comprised of 62% white/European-American and 38% Asian-American. The novice and expert group were both predominantly female. In the novice group, 69% of participants were female and 31% were male; in the expert group, 88% of participants were female and 12% were male.

Data collection and analysis
Pre-test metrics
All data were collected in a private laboratory setting during a single meeting that lasted approximately 1 h. All participants first completed a pre-test assessment of their NOS understanding and epistemological beliefs about science (descriptions of these metrics and analysis are described below). Participants spent approximately 20 min completing the pre-test assessment. The total pre-test with the scientific epistemic beliefs (SEB) survey in tandem with the NOS items was found to be highly reliable (28 items; $\alpha = .85$). Due to our small sample size and the total number of items on the pre-test, we were unable to perform factor analysis as a test of validity. However, the validity of the initial assessment tools, slightly modified for this study, was verified (Lederman et al. 2002; Lederman et al. 2014), which was critical to drawing meaningful and useful statistical inferences (Creswell 2014). The NOS assessment included items relevant to inquiry that were originally published as part of the Views of Nature of Science (VNOS) (Lederman et al. 2002) and Views About Science Inquiry (VASI) (Lederman et al. 2014) assessments, and one item that is unique to this study to assess what students think about inquiry versus science in general (Table 1).

We opted not to include the full VNOS or VASI because some aspects (e.g., difference between a theory and a law in VNOS and difference between scientific data and evidence in VASI) were not relevant to our study. Furthermore, given the open-ended nature of the questions, survey fatigue was a concern. We chose VASI questions for 1A and 1B because they were designed to assess the user's understanding

of a lack of a scientific method and that research studies begin with a question, but do not necessarily test a hypothesis (Lederman et al. 2014). Questions 2, 3, and 4, from the VNOS, assessed what the participants knew about the generation of scientific knowledge. Understanding the variety of ways one can generate scientific knowledge may reflect an understanding that there are many ways to justify scientific information and no single universal way for generating science knowledge. The final question, from the VNOS, which dealt with whether theories change, was chosen to assess what participants understood about the certainty of scientific information, and if they understood that scientific information can change in light of new evidence. Based on the questions chosen, and their relationship to epistemological beliefs about science, we coded all open-ended responses based on two NOS principles: the understanding of the lack of a universal scientific method (principle 1) and the understanding of the tenuous nature of science (principle 2). Data were coded blind by two independent coders, and overall agreement was 60% agreement for principle 1 and 68% for principle 2. Disagreements were settled through mutual discussion.

We used the SEB survey (Conley et al. 2004) to assess participant's epistemology. Although this metric was initially designed for use with elementary school students it has been used with older age groups including high school (Tsai et al. 2011) and undergraduate (Yang et al. 2016) students. The SEB survey included 26 separate items divided into four dimensions: source, justification, development, and certainty. Each dimension was counterbalanced. Participants were asked to rate each item on a 5-point Likert scale from strongly disagree to strongly agree. The source and certainty dimensions were reverse coded. Higher scores indicated more sophisticated scientific epistemic beliefs. For each domain of SEB, individual items were averaged together to create a single score for each of the four domains. Total score on the scientific epistemic beliefs survey was found by calculating the average across all items.

SCI simulation module
After completing the pre-test, all participants were logged into the *Unusual Mortality Events* SCI simulation. Novices completed the simulation in an average of 29.9 min ($SD = 15.5$) minutes, whereas experts completed the SCI simulation module in an average of 48.5 min ($SD = 17.02$). SCI simulations provide a simulated, authentic science inquiry experience within the confines of a typical classroom (Peffer et al. 2015). Although SCI is only a simulated version of authentic science inquiry, it maintains many of the features of authentic inquiry. For example, SCI simulations use real-world data, allow users to generate independent research questions, facilitate non-linear investigations with multiple opportunities to revise

Table 1 Nature of Science items included in this study

Item	Source
A person interested in birds looked at hundreds of different types of birds who eat different types of food. He noticed birds that eat similar types of food, tended to have similar shaped beaks. For example, birds that eat hard shelled nuts have short, strong beaks, and birds that eat insects have long, slim beaks. He wondered if the shape of a bird's beak was related to the type of food the bird eats and he began to collect data to answer that question. He concluded that there is a relationship between beak shape and the type of food birds eat. Do you consider this person's investigation to be scientific?	VASI; Lederman et al. 2014
A person interested in birds looked at hundreds of different types of birds who eat different types of food. He noticed birds that eat similar types of food, tended to have similar shaped beaks. For example, birds that eat hard shelled nuts have short, strong beaks, and birds that eat insects have long, slim beaks. He wondered if the shape of a bird's beak was related to the type of food the bird eats and he began to collect data to answer that question. He concluded that there is a relationship between beak shape and the type of food birds eat. Do you consider this person's investigation to be an experiment?	VASI; Lederman et al. 2014
What, in your view, is science? What makes science (or a scientific discipline such as physics, biology, etc.) different from other disciplines of inquiry (e.g., religion, philosophy)?	VNOS; Lederman et al. 2002
What, in your view, does it mean to practice science? In other words, what do you think scientists do when they engage in research?	This study
Scientists perform experiments/investigations when trying to find answers to the questions they put forth. Do scientists use their creativity and imagination during investigations?	VNOS; Lederman et al. 2002
After scientists have developed a scientific theory (e.g., atomic theory, evolution theory), does the theory ever change?	VNOS; Lederman et al. 2002

hypotheses, including dead end or anomalous data, and engaging participants in the process of doing science (Chinn and Malhotra 2002; Rowland et al. 2016).

The version of *Unusual Mortality Simulation* used here was a modified version of the module used in Peffer et al. (2015). Changes included additional freedom to perform actions and updated information in the in-simulation library. These changes were made to allow for autonomy while participants engaged with the simulation and to update content with current scientific knowledge around Unusual Mortality Events in the Indian River Lagoon. The SCI simulation web application captured data as the participant completed the simulation including the order in which actions (such as tests, generation of new hypotheses) were performed, as well as the participant's responses to open-ended questions embedded throughout the module to interrogate why participants performed certain actions. Additional information about each participants' practices was captured through screen-capture recordings made during the module. As participants completed the simulation, they verbalized their thought processes using a think-aloud procedure (Someren et al. 1994) and these recorded thought processes were later transcribed. Upon completion of the simulation, participants completed a demographics survey and were immediately interviewed about their strategy and rationale for certain decisions.

Mixed-methods analysis
We used a convergent parallel mixed methods design (Creswell 2014) to assess differences in expert and novice practices in authentic science inquiry, and their

relationship to epistemological beliefs. In this type of mixed-methods research, quantitative and qualitative data are collected simultaneously and analyses are combined to yield a comprehensive analysis of the research question(s) at hand (Creswell 2014). We felt that mixed-methods research was best suited to answering our research questions because fully distinguishing the practices of experts and novices required a qualitative approach, but comparing scores on pre-test metrics and relating these scores to practices required a quantitative approach. Obtaining different but complementary qualitative and quantitative data not only allows for a greater understanding of the research problem, but it also enables researchers to use the qualitative data to explore the quantitative findings (Creswell and Plano 2011).

Quantitative analysis
We first counted the total number, and type, of actions performed by each user. To determine the number and type of actions participants made, the lead author reviewed each of the screen-capture videos and the logs created by the SCI simulation engine. Actions were categorized as either investigative or information seeking. Investigative actions included the generation of a new hypothesis, performing a test, or making conclusions. These investigative actions are aligned with models of scientific activity that include experimentation, hypothesis generation, and evidence evaluation (Osborne 2014a). How evidence is evaluated may provide insight into epistemological beliefs such as certainty. For example, a student who believes scientific information to be unchanging may include very little evidence or few

tests since there would be nothing else to examine once a final answer is reached. Information seeking actions were any time the user sought additional information as part of his or her inquiry process, including use of internet search engines or various features built into the simulation such as the library, lab notebook, and external links within the simulation. Information seeking actions could be considered as part of the process of experimenting, since users were seeking information about the outside world. However, we chose to distinguish using outside information as a different action within the simulation since it is a known engineering practice chosen by experts (and not novices) in authentic situations (Atman et al. 2007). Even though engineering and science practices are not identical, we still felt that pursuit of information as part of a project was likely analogous between the two disciplines. Whether or not a student chooses to seek outside information and what information is sought could provide insight into a participant's epistemological beliefs about source. For example, is the simulation the ultimate authority, or are there other valid information? What kind of information is sought, peer-reviewed literature or news articles for the general public?

Each investigation was coded blind by two people as simple or complex in nature. The primary author served as the tie-breaker for disagreements. Based on the type of data and number of raters, Cohen's kappa was used to assess inter-rater reliability. A Cohen's kappa of 0.533 (p = .003) indicated a moderate level of agreement among raters. A simple investigation is reminiscent of the simple inquiry described by Chinn and Malhotra (2002), where the user performs a few tests until a basic cause and effect relationship is uncovered, at which point they make their conclusions. A complex investigation is one in which the user performs a multi-pronged investigation with multiple cause and effect relationships. The user may perform many related tests with the goal of developing a model that describes some kind of underlying mechanism. The user seeks to connect different sources of evidence in a manner to explain how they are both related to each other and the problem at hand. For example, a participant may choose to relate algal blooms to explain the lack of seagrass, which would explain why a foreign substance was found in the stomach of manatees instead of the sea grass, which is their normal diet. In contrast, a simple investigation would note that sea grass was dead and conclude that as the main cause of the unusual mortality events without any attempt at explaining why sea grass was the cause, or how it related to other lines of evidence. A more complex investigation with multiple lines of evidence may indicate underlying epistemological beliefs about certainty, since there are multiple causes and not a single correct answer. A

subset of the data was analyzed for linguistic features to determine if novices or experts differed in the type of language used during the concluding phase of their investigations. These results are reported elsewhere (Peffer and Kyle 2017), but we include the findings here as a variable—expert verb score—and therefore as part of our model to describe expert versus novice practices.

Statistical analysis

Quantitative analysis of the data was performed using SAS 9.4 (SAS Institute Inc 2014). To determine differences between expert and novice performance on the pre-test, we used Fisher's exact test instead of chi-square test due to small counts for some categories of certain variables. To predict total number of actions, as well as total number of investigative and information seeking actions, we fitted Poisson (count) regression models with logarithmic link function. Poisson regression modeling is a member of generalized linear models used when modeling count data following a Poisson distribution. PROC GENMOD in SAS 9.4 was used to fit the Poisson regression model, using the maximum likelihood technique to estimate the parameters. The predictors used within the Poisson count regression models were SEB survey performance total average score, expert verb score, and both NOS principles. In this study, significance level was considered .05 for all tests.

Qualitative analysis

Within each category (complex or simple), we chose the average participant and made comparisons with others in the group as appropriate to form an expert dyad and novice dyad. The average participant was determined based on the temporal pattern of types of actions performed (Figs. 3 and 4). Since the order of actions was an important distinguishing factor between participants, we opted to use this as our primary criterion for determining the average participant rather than the user's number of actions (Table 2). We do note that for both dyads analyzed here, the number of information seeking actions was higher in the expert groups, which was a trend among our full data set. For consistency, all case study examples are female and assigned a pseudonym. Since the majority of novice participants were African-American, and it was not possible to pick a dyad that corresponded in race to the expert population, we decided to choose two African-American students in their senior year of college. The simple novice, Sally, majored in human learning, and development and the complex novice, Beth, majored in psychology. The simple expert, Lisa, was European-American and in her fourth year of doctoral studies in neurobiology. The complex expert, Janet, was Asian-American and in her fifth year of doctoral studies in biology.

Table 2 Summary of actions performed by all participants

Expertise	Investigative Actions	Information Seeking Actions	Total Actions	Investigation Type
Expert	24	45	69	Complex
Expert	15	34	49	Complex
Expert	9	17	26	Complex
Expert	15	7	22	Complex
Expert	5	8	13	Complex
Expert	14	20	34	Simple
Expert*	25	5	30	Simple
Expert	9	2	11	Simple
Novice	18	11	29	Complex
Novice	13	12	25	Complex
Novice	11	11	22	Complex
Novice	19	3	22	Complex
Novice	15	3	18	Complex
Novice	7	9	16	Complex
Novice	7	2	9	Complex
Novice	7	32	39	Simple
Novice	13	15	28	Simple
Novice	12	12	24	Simple
Novice	5	18	23	Simple
Novice	8	11	19	Simple
Novice	7	10	17	Simple
Novice	7	8	15	Simple
Novice	9	0	9	Simple
Novice	7	1	8	Simple
Novice	7	1	8	Simple
Novice	7	0	7	Simple
Novice**	5	0	5	Simple
Novice	3	0	3	Simple

Highlighted rows indicate participants chosen for case study analysis

The qualitative analysis focused on two sources of information: the logs created as each participant engaged in the simulation (the participant's lab notebook) and the think-aloud transcripts. These two sources of data were analyzed for features that were indicative of authentic science inquiry and/or expertise, such as searching for outside information, which is an expert practice in engineering and may be similar in authentic science inquiry (Atman et al. 2007) or the use of tentative language (Peffer and Kyle

2017). To ensure trustworthiness of our qualitative study (equivalent to reliability and validity of quantitative studies), we used triangulation between the qualitative information, and the quantitative metrics such as how many of each type of actions was performed. We also included rich descriptions in this manuscript to allow readers to form their own conclusions, and involved an external auditor who reviewed this project and indicated her agreement with the conclusions presented here.

Results

Experts perform better on established metrics of nature of science, but not epistemology

First, differences in expert and novice populations were assessed using previously established metrics of NOS and science epistemic beliefs. Experts scored higher than novices on both nature of science principles assessed (Fig. 2). For principle 1, 18 novices had a naïve score while seven experts had a mixed or sophisticated score. Only one expert scored as naïve and two novices scored as sophisticated. For principle 2, 12 novices scored as naïve and eight scored as mixed or sophisticated, while all eight experts scored as mixed or sophisticated. The small counts per some combinations of different categories of factors prevented us from performing the chi-square tests of association; instead, the Fisher's exact test, which is as powerful, was conducted to check whether these observed associations were statistically significant or not. A two by two contingency table was created for both nature of science principles by aggregating the mixed and sophisticated scores to test association between expertise and NOS principle scores. Association between expertise and both principles 1 and 2 scores were statistically significant, using Fisher's exact test (p = .0002 and p = .0084, respectively).

No associations were observed between expertise and scores on the science epistemic beliefs metric on any of the four domains assessed (justification, certainty, development, and source) nor on the total aggregated scores (Table 3). Experts scored marginally higher on the source and certainty domains and the overall total score. However, scores among all participants were very high, mostly above four on a five-point scale.

Experts and novices have distinct practices in authentic science inquiry

Investigative style and general patterns of actions (Figs. 3 and 4) performed during authentic science inquiry were assessed to examine differences between experts and novices. Investigative style was separated into two categories: simple and complex. Experts perform more complex investigations than novices (62.5% versus 35%, respectively), and novices perform more simple investigations than experts (65% versus 37.5%, respectively).

General pattern of actions between novices and experts, specifically if the number of investigative, information seeking, and total number also, was investigated. Complex novices performed more investigative actions (M = 12.86, SD = 4.85) than simple novices (M = 7.46, SD = 2.70). In contrast, complex experts performed less investigative actions than simple experts (M = 13.60, SD = 7.20 and M = 16.00, SD = 8.19, respectively). Overall, we observed that experts performed more investigative actions than novices (M = 14.50, SD = 7.09 and M = 9.35, SD = 4.36, respectively). For information seeking actions, complex novices performed slightly fewer actions than simple novices (M = 7.29, SD = 4.42 and M = 8.31, SD = 9.60, respectively). Complex experts on average performed over twice as many information seeking actions as simple experts (M = 22.20, SD = 16.72 and M = 9.00, SD = 9.64, respectively), and the simple experts performed slightly more information seeking actions than either of the novice groups. Experts overall sought more information than novices (M = 17.25, SD = 15.27 and M = 7.95, SD = 8.04, respectively).

The quantitative part of the analysis complemented our observations regarding the relationship that existed between the category of expertise participants belonged to (expert or novice) and the number of actions they
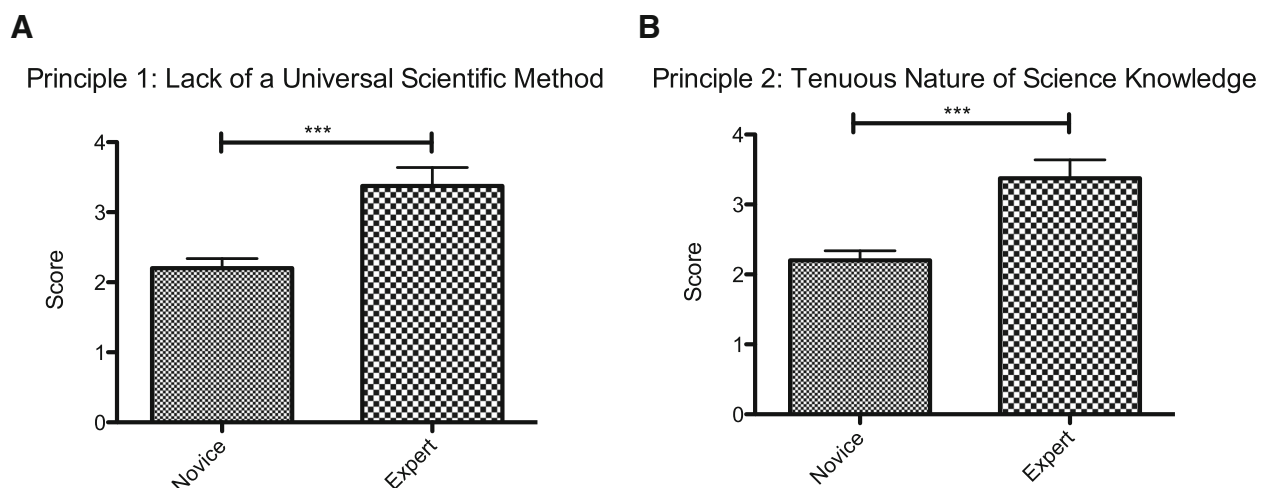

Fig. 2 Experts performed better than novices on a pre-test assessment of their NOS knowledge on both principle 1, lack of a universal scientific method (**a**) and principle 2, tenuous nature of science knowledge (**b**)

Table 3 Average and standard deviations for novice and expert scores on the scientific epistemic beliefs survey

Dimension	Novice		Expert	
	M	SD	M	SD
Source	3.87	0.88	4.35	0.40
Justification	4.31	0.39	4.33	0.37
Development	4.38	0.46	4.38	0.43
Certainty	4.06	0.69	4.46	0.33
Total	4.18	0.37	4.38	0.28

performed. Fitting a logistic regression model, using PROC LOGISTIC within SAS 9.4, it was shown that the total number of actions performed significantly predicted whether participants belonged to the expert or novice category of expertise (*Likelihood ratio* χ^2 = 6.22, *p* = .013; *Wald* χ^2 = 3.95, *p* = .047). The Hosmer and Lemeshow goodness of fit test detected no evidence of lack of fit of the aforementioned logistic regression model (*p* = .19), implying that the design was a good fit.

Actions in context were examined by using randomly selected videos to generate profiles describing the sequence of inquiry events (e.g., information seeking, hypothesis generation) for five simple and five complex novices (Fig. 3) and three simple and three complex experts (Fig. 4). Among both experts and novices, as more actions were performed, the frequency of information seeking actions increased. Among all experts analyzed,

the complex novices, and simple novices N4 and N5, demonstrated an iterative process of moving back and forth between investigative and information seeking actions. These informative seeking phases were remarkably long, particularly in complex experts, E5 and E6 (Fig. 4). Both E5 and E6, in addition to N10, began their investigations with seeking information and continue to look for outside information regularly throughout their investigation. Among both simple novices and experts, there was a trend towards short periods of information seeking, when one external resource was utilized, rather than an in-depth review of literature. We now turn to an in-depth discussion of practices by four representative participants.

Case studies

The simple novice: Sally Sally (N3, Fig. 3) generated two hypotheses, performed three tests, and did one information seeking action. Sally began her investigation by generating two hypotheses, "Could the cause of the deaths be due to contamination that has taken place in the lagoon? Are the animals who live in the lagoon killing each other off due to lack of food" with the rationale, "…because pollution leads to contamination and it usually has a lot to do with spikes in the death of animals… because I learned that there are many different animals that live in the lagoon that may be lacking food to eat

Fig. 3 Inquiry trajectories of novice participants

Fig. 4 Inquiry trajectories of expert participants
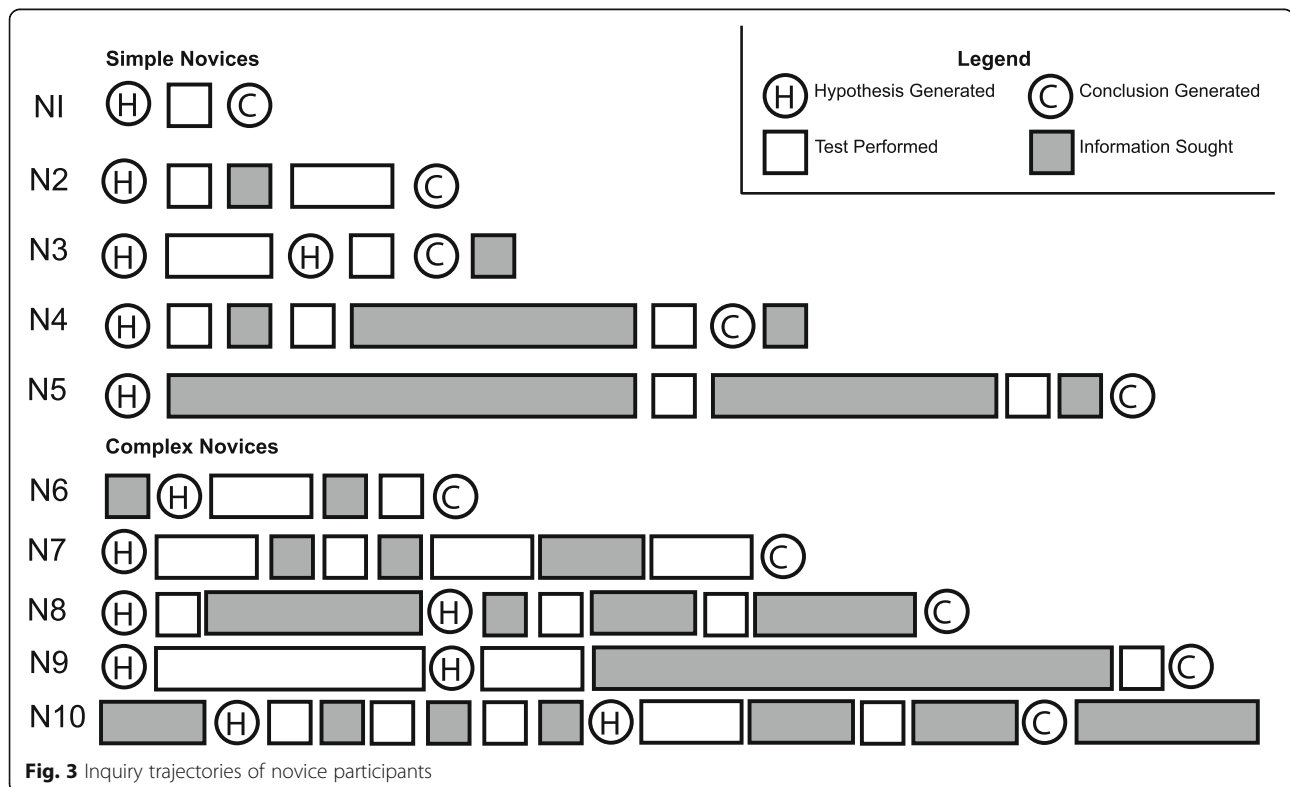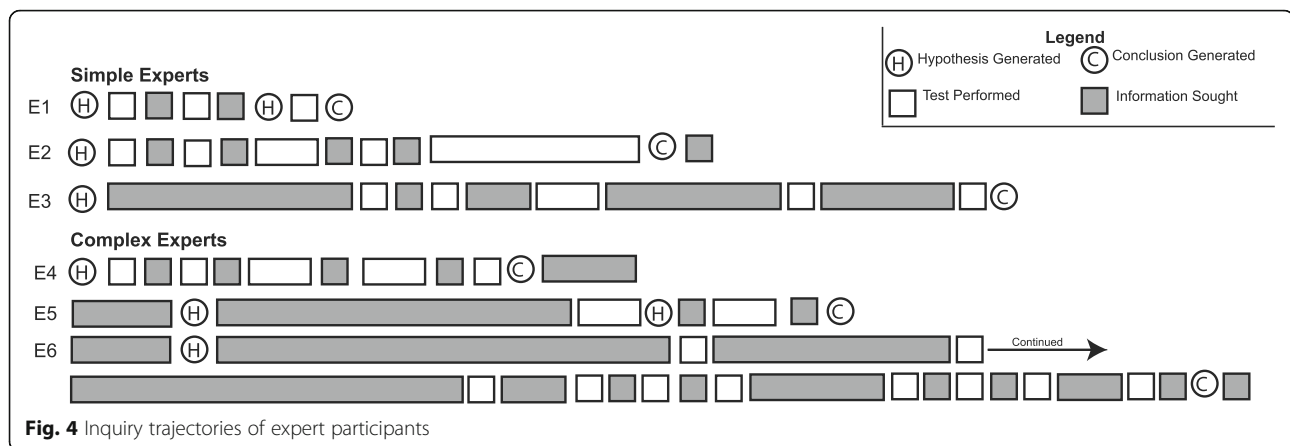
and are not surviving by eating other animals." Sally then stated in response to the query "How would you like to test your hypothesis?" that she "would like to go into the lagoon and test the waters for contamination, and watch the animals' behaviors." In her think-aloud transcript, Sally mentioned that she wanted "to come up with a hypothesis, but [she] was not sure of the background information [and wondered if] she should guess at one?" This is a behavior identified as characteristic of novices, rather than pursue additional information in light of what they do not know, they move forward with their investigation. In contrast, we see that N6 and N10 (Fig. 3) and E5 and E6 (Fig. 4), all coded as complex, began their investigations with more effort put forth in gathering information before executing their first test.

After completing her hypothesis generation phase, Sally then performed a test to check water salinity in the lagoon to see if it "is at a level for animals to survive" and noted that she did not think it was a contributing factor. Sally then chose to examine invasive species in the Indian River Lagoon to see if "invasive species could be the cause." After reviewing the test results (that the Indian River Lagoon contains over 240 non-native species), she wrote in her lab notebook that yes, they were having an impact, "because they can affect how the ecosystem works in the lagoon." In her think-aloud transcript, Sally said:

"So, maybe the exotic species could be bringing in maybe some type of disease or something that the native species cannot fight off. Okay, I think I'm ready to make a conclusion, although I don't know if I should -- I'm trying to think, should I change my hypothesis? Because now I'm interested to know that if the species that arrived there from all these different places, or if they could have maybe caused the threat to the ecosystem which now has spiked the mortality rate, or the mortality rate has gone up in the native species that were there."

Sally decided to generate a new hypothesis, rather than make a conclusion. Her new hypothesis was "Are the species that are brought into the lagoon causing the ecosystem to change which is causing the mortality rate to increase?" with the rationale, "The fact that I found new information about other species migrating to the lagoon and this could affect the mortality rate." Sally stated that she wanted to test her hypothesis "By seeing how the ecosystem has changed since the migration of the new species, and how the changes could affect the native species." What is unique to Sally's investigation is that after generating this hypothesis, she then repeated the same invasive species test, because she had "learned that the invasive species can cause change in the ecosystem." Given the design of the simulation, she received the same information again, but rather than focus on the fact that invasive species were present in the Indian River Lagoon, she instead focused on the last sentence relating to the impact of invasive species on ecosystems. She then stated that "Yes, I believe they do contribute because they can change the ecosystem which could cause the native species to not adjust to the change but die."

Sally then concluded "...invasive species...caused change to the ecosystem and now the native species cannot adapt...so they are now dying" – however, she never collected any information directly tying the presence of exotic species to the high mortality of dolphins, manatees, and pelicans in 2013. When asked during her interview how she knew she was ready to conclude, Sally said "because she had enough information." Sally's only information seeking action occurred during her conclusion phase when she checked her original hypotheses in her notebook; this lack of information seeking actions was characteristic of novices.

The complex novice: Beth Beth (Fig. 3, N8) generated two hypotheses, performed three tests and did eleven information seeking actions. Beth also had some experience working as a laboratory research assistant, but had

never presented a poster describing her own independent research nor was listed as an author on a primary research manuscript. Prior to generating her first hypothesis, she focused on the preliminary data provided by the simulation in the introduction section. In particular, she focused on the temporal distribution of dolphin deaths to devise her first hypothesis. In her think-aloud transcript she stated:

"[Dolphin stranding] spikes in March…it's like spring, summer. Maybe people are on the boats…I'm thinking maybe tourist/people attracted to the area, and possibly boating accidents…Okay, my hypothesis: the unusual increase of deaths of dolphins in this area could be attributed to tourist/boating population in the area…looking at spring/summer months in which deaths increase, and in which are known to be times of the year where tourists/boating is more popular. How would I like to test it? I would like to examine tourist/boating rates during the same time period in which these deaths are occurring."

Beth's first test is to examine dolphin necropsy[1] results because she wanted "to see how they died…possibly trauma from boating accidents?" in her think-aloud transcript, she notes that the necropsy report indicated that dolphins were emaciated. She says "what does that mean?" and proceeded to use Google to look up the definition of emaciated. After looking up the definition, she then stated out loud, "Thin or weak. Lack of food, so that has nothing to do with that," likely referring to her original hypothesis. She then went on to say "No evidence of entanglement. Skin, eyes, mouth generally normal. Some animals have skin lesions. Oh, brain infections, that's bad…No presence of any dangerous toxic poisoning…15 out of 144 was positive [for morbillivirus]. What is morbillivirus?" Beth then followed a simulation link that provided additional information about morbillivirus and then used Google to search the phrase "How do you get morbillivirus" and from there followed a link for a NOAA.gov website. After pursuing this information, Beth then said, "My hypothesis is definitely refuted. No evidence of head trauma as would be seen in boating accidents. I want to generate a new hypothesis." What is notable about this statement is that Beth says that her hypothesis is *refuted* not wrong or incorrect. Our previous work (Peffer and Kyle 2017) examining the language used by experts versus novices suggested that experts use more hedging language, such as refuted or supported. Notably, here we observe a novice with a more complex style of investigation using more expert like language.

Beth then generated a second hypothesis, "The spread of morbillivirus is causing the unusual high rates of death among dolphins." Her next test was to examine manatee necropsies because they are "closer in relation/environment to the dolphin." After reviewing the results and determining that the manatees had inflamed gastrointestinal tracts and red algae in their stomachs, she read more about the algae and stated that her hypothesis was "Supported. The discovery of [red algae] could be a possible method of transmission amongst and between dolphins/manatees/other sea life in procuring morbillivirus." At the transition page where the participants can choose to do another test, generate a new hypothesis, or make their conclusions, Beth said "All right. I want to do one more test just to make sure. Okay, they said it was algae, right? It's algae? Okay, so I'm going to [do the] algal blooms [test]." After reviewing the test results and seeking additional information on algal blooms, she stated "Yes. [algal blooms are contributing to the unusual morality events] Algal blooms are contributing to toxins which in turn can possibly manifest into either this virus or the death of surrounding sea life."

During the conclusion phase, Beth returns to the morbillivirus resource to determine what other animals are at risk for morbillivirus. She then went on to say "But how do you get [morbillivirus] in the first place? Must be like a reaction to something, which may be that algae. Okay, my final conclusion is the high rates of deaths amongst dolphins can be attributed to morbillivirus, and which may be surmounting due to a reaction to algal blooms." When asked to provide evidence that supported her conclusion, she stated, "Autopsies, Algal bloom rates, Virus information concerning effects on the body in which seen in the autopsies," but gave no evidence for how morbillivirus could be related to increased numbers of algal blooms.

Summary: Comparison of simple and complex novices Both novice participants performed the same number of investigative decisions, and were slightly below the average number of investigative actions performed by novices. The complex novice (Beth) performed significantly more information seeking actions than the simple novice (Sally). We also noted that complex novices left the simulation to pursue other information. Simple novices, if they performed information seeking actions, typically only used sources linked from the simulation itself. Both Sally and Beth demonstrated a logical structure to their investigations, but the complex novice, Beth, was more detail oriented and her investigation followed a logical pattern. Rather than stating "I don't know" like Sally, Beth reviewed the information until she felt that she had a good starting point for her investigation. Beth used more expert like language as she thought aloud using terms such as "supported" and "refuted" whereas novices generally used either less

expert language (e.g., "my hypothesis is wrong"). All participants were proficient English speakers and previous work (Peffer and Kyle 2017) indicated that written use of tentative language was a hallmark of expert practices. With both of the novice participants, their conclusions were ill-supported by the evidence collected. In Sally's investigation, she concluded that the invasive species were contributing to the deaths of dolphins, manatees, and pelicans because they are generally known to disrupt ecosystems. In Beth's investigation, she proposed a link between morbillivirus and algal blooms, but gave no evidence to support this idea other than both instances were occurring at the same time.

The simple expert: Lisa Lisa (E1, Fig. 4) generated two hypotheses, performed three tests, and did two information seeking actions. We also noted that Lisa had the simplest investigation of all of the simple experts, but since the average simple expert was male, for consistency purposes we chose Lisa. Similar to Sally, who spent seven minutes reviewing the introductory material, but not Beth, who spent three minutes, Lisa spent ten minutes, the longest of all four participants highlighted in these case studies, reading and thinking-aloud about the introductory material, including taking notes by hand. When initially asked as to what she thought was causing the animal deaths, Lisa entered into her notebook, "I don't think I have enough information at this point to hypothesize the cause of the UMEs," and stated out loud, "I'm not sure I can really postulate with this much information. Perhaps I'm missing something." When prompted to make a hypothesis, Lisa stated that her hypothesis was "In 2013, human interactions increased the number of unusual mortality events in manatees, dolphins, and pelicans in the Indian River Lagoon." In her think-aloud transcript Lisa stated that she "was always taught to phrase [her] hypothesis very, very carefully." When asked how she would like to test her hypothesis, Lisa stated, "I would need access to data reporting human interactions in the Indian River Lagoon in 2013. I would also need to identify a similar ecosystem (or several) in another part of the world with differing levels of human interaction but housing the same species. I would compare UMEs for these species in the Indian River Lagoon to the comparable estuaries elsewhere. This would not allow me to infer causation but could yield correlational information." There are two notable observations from this statement. First, Lisa is discussing the need for a control group. Although control of variables strategy is identified as an important developmental milestone as children learn to think scientifically (Schwichow et al. 2016), it can also be indicative of a less sophisticated epistemology because it may represent an understanding that science inquiry can

only proceed with controlled experiments, whereas some disciplines of inquiry rely instead on observational data. The second observation is Lisa's concern with causation and correlation. She is acknowledging the importance of identifying a mechanistic link between two observed phenomena, which may point towards a more sophisticated epistemological stance. We also note that in her think-aloud transcript at this point in her investigation that Lisa states "I'm not sure with causation how you would do that. You can't put it in a lab." which may indicate some acknowledgement that there are a variety of scientific practices. However, the focus of lab work may indicate the presence of a less sophisticated belief that all science occurs in labs.

After generating her hypothesis, Lisa then examined dead affected dolphins, because this "could provide information about whether or not an infectious disease or biotoxin is implicated in the UMEs." While reviewing the findings, Lisa follows a link embedded in the simulation to read more about morbillivirus. She also noted aloud that "15 of the 144 [examined were positive for morbillivirus], which is not conclusive." When reflecting upon these results, Lisa noted in her lab notebook that the "Most striking…findings [were] that animals were emaciated, and also that the conditions they present with [were] consistent with a morbillivirus infection, yet the majority of the dolphins did not test positive for this type of infection. My hypothesis is neither supported nor refuted by this information." Of importance here is her use of tentative language, and that she notes that only some dolphins are sick; interestingly, her reflection does not fully explain the wide scale of the 2013–2014 unusual mortality events.

Lisa next tests to see if there are invasive species present in the Indian River Lagoon. After learning that there are many different invasive species in the Indian River Lagoon and following an embedded link within the simulation to read more about invasive species, she states in her notebook that she "[does not] think [that invasive species are contributing to the unusual mortality events]. The identified invasive species are things such as sea squirts and mussels and I would not expect these species to greatly impact the health of dolphins and manatees…On the other hand, the presence of invasive species could limit the food supply for the larger species being impacted by UMEs so perhaps it is a contributing factor." Lisa then decides to generate a new hypothesis stating aloud:

"we had global warming[2] with the invasive species. And then we have a virus and dead dolphins. I'm getting to a point where I feel like it may not be just one of these divisions. It may not be just infectious diseases or just ecological factors. It may be that

global warming has brought in these invasive species, and also maybe some biotoxins or infectious diseases have been allowed to propagate."

Lisa enters her new hypothesis into her lab notebook as "In 2013, climate change contributed to an increase in UMEs in dolphins, manatees, and pelicans in the Indian River Lagoon," giving the rationale that "...the information that I recently accessed implicating that climate change encouraged the intrusion of invasive species into this particular ecosystem. These invasive species could limit the food supply for the larger mammals and warmer temperatures could also support growth of biotoxins that might cause UMEs," and stating she would like to test her hypothesis by "Compar[ing] the UMEs in this estuary with the UMEs in an estuary that has colder temperature water. If they differ, look specifically at the level and types of invasive species and biotoxins." Lisa's desire to use a comparison group as part of testing her hypothesis and her saying there may be two possible causative factors, not a simple cause and effect relationship, were notable since use of a control group can indicate either a sophisticated understanding of scientific practices, namely the importance of a control, or a less sophisticated understanding indicating that all scientific investigations require controls.

Lisa performed one additional test, examining the water temperature, because her "hypothesis features warmer water as a possible cause of UMEs." After determining that the median water temperature was at the high end of normal in 2013, she states in her lab notebook, "Yes, I think the water temperature is contributing to the unusual mortality events because when a median of 25.09 degrees C is reported, this means there were much higher temperatures recorded throughout the year as well. The average temperature range is up to 25 degrees C, which does not diverge from the 2013 reported median temp, but average and median are different types of statistical data." Lisa's final conclusion was "Global warming played a role in the increase in UMEs in 2013," and she cited "increased water temperature and the enhanced presence of an invasive species (sea squirts) that has been shown to migrate to warmer temperature waters" as evidence for her conclusion.

The complex expert: Janet Janet generated two hypotheses, performed three tests, and 16 information seeking actions. Similar to Lisa and Beth, Janet also spends time gathering information at the start of her investigation. Like Lisa, Janet also requested paper to help organize her thoughts. Janet is unique in that she began to look for additional information at the very beginning of the simulation even before generating a hypothesis. This was a behavior generally characteristic of the complex

experts (Fig. 4) and consistent with novice/expert studies in the engineering education literature (Atman et al. 2007). After reviewing the background material in the simulation, Janet used Google to search "estuaries and undetermined deaths of animals." After the results appeared, Janet said aloud "Okay, so a lot of stuff popped up. Which one will I go to first? Maybe going to the New Yorker because it seems like it's a reliable, quick read source." This comment by Janet is notable as this was one of the few instances where a participant discussed the reliability of the source, which is another facet of epistemology (Barzilai and Zohar 2012; Mason et al. 2011). As Janet was developing her hypothesis, she continued to seek information and in her think aloud transcript discussed different sources that she would use to find information. Janet says:

"I do think it's pollution, but I don't know where -- actually, I should be going onto PubMed[3] instead of Google now...so typed in estuary and animal deaths into PubMed... Nothing too relevant...now that I typed in lagoon instead of estuary, one specific article came up...before I want to write a hypothesis, I want to make sure I have at least a decent hypothesis, so I'm going to look at some more relevant articles hopefully on PubMed, and see what else they say. The first one said that it was due to cyanobacteria changing the metabolism of the animals in the Brazil Lagoon...Okay, so now I've gone back to Google because I want to see...if there's, like, a general broad view, like different type of perspectives on what could be the cause. Then that will help me try to consider all the possibilities that was involved"

While generating her hypothesis, Janet switched between primary literature found in PubMed and general information articles found in Google. From her transcript, there appeared to be some rationale between her decisions to search for information in each location. In general, experts would seek out information in the primary literature, but novices never looked in the primary literature. This may provide some insight into Janet's epistemological beliefs as they related to the source of scientific information, namely the use of primary literature. When the investigator queried Janet's decision to refer to the same article repeatedly while generating her hypothesis, Janet stated aloud "[the article] at least cited some science or something talking about this. And it looks like it has some scientific backing. And what the point is seems to be decent in that it's not too [irrational]."

Janet's hypothesis, "There is high death rates in the lagoon, because the environment of the lagoon is not stable, that is temperature/pH/salinity/02 levels have

changed," and her rationale for this hypothesis, "From observing the pie chart and it showing what factors accounted for the deaths," referred to information collected at the beginning of her investigation. When asked how she would like to test her hypothesis, Janet stated in her lab notebook, "By creating a pseudo lagoon in the lab then altering each of those factors and seeing how it will affect the fish/animals that will be put in there." Again, like Lisa, we see the effort to create an experimental setting that will allow for control of variables.

Janet spent the bulk of her investigation in this preparatory, information seeking mode (eight minutes) and the most time of all four participants examined in these case studies generating her first hypothesis (fourteen minutes). After generating her first hypothesis, Janet tested the average dissolved oxygen in the Indian River Lagoon "because it is one of the independent variables that I think is contributing to throwing the lagoon off. Maybe O2 levels are lower, causing sickness." Notably, the term "independent variable" was not utilized by novice participants. After reviewing the results, Janet stated that dissolve oxygen did not contribute to the unusual mortality events. Janet then tested water salinity stating, "Salinity is an independent variable that is potentially contributing to the deaths, and lagoons are kind of where fresh and salt water meet, so maybe an imbalance will cause problems." In response to the data, Janet stated that there was not enough information to decide whether or not water salinity is contributing to the unusual mortality events.

Janet then decided to generate a new hypothesis, "Cyanobacteria is changing the metabolism of animals in the lagoon to make them sick," with the rationale that "There were many PubMed articles about this." Again, Janet explicitly connected her hypotheses to her collected data. Janet then stated that she would like to test her hypothesis by "Test[ing] the levels of cyanobacteria in different lagoons that had deaths and no deaths. Field experiment." We see here the emphasis on having a control group, but also on conducting a field experiment, rather than trying to conduct a field experiment in a lab which is what Lisa suggested. Janet also stated aloud "...different from my original hypothesis, I had suggested to stimulate or create a fake lagoon in the lab. I thought that would be better to control. But since this is more like measurements wise because your interest is just looking at the cyanobacteria, you can actually maybe test the tissues from the dead animals. This could be a field experiment. You don't have to do much in the lab maybe." When the investigator queried Janet as to why she decided to generate a new hypothesis, Janet said that her previous tests had "Failed… so, I'm going back to the cyanobacteria. I originally didn't go there because I had grouped cyanobacteria with biotoxin, so I thought that was already accounted for. But there was a lot of

PubMed articles about this, so I'm sure it's maybe more reliable than my hypothesis [based on an NPR article]."

Janet then used Google to determine if cyanobacteria are the same thing as an algal bloom. She described the results from Google aloud stating "I typed in cyanobacteria and algal bloom, and a lot of things popped up. For example, the first thing is from the EPA. And it says that cyanobacteria are harmful algal blooms, so they're the same thing. So, I'm going to choose look for algal bloom levels instead." From here, Janet chose to examine the presence of algal blooms in the Indian River Lagoon. She stated she wanted to do this test because "Because articles from PubMed suggest that cyanobacteria is the culprit for deaths in the lagoon and algal bloom is cyanobacteria." Janet then followed links embedded in the simulation, including an article on Brown Tides and a recording of historical harmful algal blooms in the Indian River Lagoon. She concluded aloud that algal blooms were a contributing factor "because they can block sunlight and deplete oxygen." From here, Janet said, "I wanted to do one more test because looking at algal blooms is not enough, but that option isn't available. So, I'm going to click on I'm ready to make my conclusion." In her notebook, Janet listed her final conclusion as, "Cyanobacteria like algal blooms are contributing to deaths perhaps by decreasing O2 levels and sunlight," and the evidence she collected to support this conclusion as, "The damaging history of algal blooms and PubMed articles."

In Janet's post-simulation interview, she made several epistemologically relevant comments. For example, when asked if this simulation changed her perceptions of science, she stated, "No, but it did bring to my attention that I probably am not too confident about the different types of experiments. Because what we do in the lab is usually hypothesis…driven. And so, like, this is kind of hypothesis, but it's more about observation. There's this issue with the lagoon. Like, what do you think is happening? And then [a question on the pre-test queried if a scenario was] an experiment? So, I realized I was kind of fuzzy on that, and I wasn't sure if there's a different type of experiment that is not like one where you have to test, where you can just observe." In response to why she decided to conclude, Janet stated that she "really wanted to get it right. Not that you can get it right." Although there may be some ambiguity on what the participant meant is "right" (a correct answer, versus a proper experimental design), it is notable that she then clarified her comment by saying that it is not necessarily possible to get something "right," which could indicate a more sophisticated epistemological stance.

Summary: Comparison of simple and complex experts Lisa and Janet were similar in that they spent

considerable care and effort taking notes and exploring the background information before fully engaging in the simulation. However, Janet pursued extensive amounts of additional information prior to generating her hypothesis. Both experts proposed experiments to test their hypotheses involving a control group. Experts may have acknowledged the importance of a control group as part of experimental design because of training that required more experimental, rather than observational work, or an indication that there is only a "single" correct way of doing science. Lisa and Janet both commented that there may be a "single," correct way of doing science, which may also result from their scientific training. Focusing on a single correct method of doing science, one that requires control groups, would indicate a less sophisticated understanding of science practices, whereas acknowledging multiple ways of doing science would be a more sophisticated practice. Janet pursued more information, to a deeper level, specifically commenting on why she was pursuing different sources, whereas Lisa tended to use only information provided by the simulation, more similar to novice behavior. Both experts followed a logical pattern, but Lisa's investigation was more cursory and ended when she obtained a test result with a plausible explanation, rather than pursuing a more complex relationship such as Janet.

Pre-test performance predicts authentic inquiry practices

Given the wide variety of practices observed in both experts and novices, putative predictive practice models based on pre-test performance were generated. First, performance on the pre-test metrics was predicted by expertise (Fig. 2) and predicted the total number of actions. The first Poisson count regression model was built to predict the total number of actions, using different predictors including demographic information of the participants,

$$\log(\mu) = -0.38 - 0.36 Gender^* - 0.46 Race_1^* + 0.05 Race_3$$

$$+ 1.07 Expertise^* + 0.78 SEBtotal^*$$

$$+ 0.06 NOSprinciple1_2 + 0.33 NOSprinciple1_3$$

$$+ 0.23 NOSprinciple2_2 + 0.41 NOSprinciple2_3^*$$

$$- 0.11 ExpertVerbScore^*,$$

where, μ is the average number of actions. Asterisks specify significant coefficients in the model; that is, the predictors that significantly contributed in the process of predicting the total number of actions performed by the participants. Table 4 shows the parameter estimates of the first Poisson count regression model for predicting the number of actions. Additional file 1: Table S1 shows the test of main effects for this model.

The results of the first Poisson regression model demonstrated that NOS principle 2 (tenuous nature of science), but not NOS principle 1 (lack of a universal scientific method), significantly predicted total number of actions. This variable had three categories of 2 (naïve), 3 (mixed), and 4 (sophisticated), where a 4 was used as the baseline category. Coefficient of 0.23 for $NOSprinciple2_2$ means, holding the other variables constant in the model, the difference in the logs of expected number of actions is expected to be 0.23 units higher for participants with naïve understanding of the tenuous nature of science, compared to those whose understanding was sophisticated. For example, an average of ten actions for the population of students with a sophisticated score in the tenuous nature of science corresponds to about 12 actions for a similar population of students with naïve scores. Coefficient of 0.41 for $NOSprinciple2_3$ means, under the same circumstances, the difference in the logs of expected number of actions is expected to be 0.41 units higher for participants with mixed understanding of the tenuous nature of science, compared to those with sophisticated understanding. Here, an average of ten actions for the population of students with a sophisticated score in the tenuous nature of science corresponds to an expectation of about 15 actions for a similar population of students with mixed scores.

When building the second Poisson count regression model to predict the number of information seeking actions using the same predictors used for the first model, NOS principle 1 (lack of a universal scientific method) ($\chi^2 = 1465$, $p = .0007$), but not NOS principle 2 (tenuous nature of science) was a significant predictor. For the third Poisson count model, which was built to predict the number of investigative actions using the same predictors as the first and second models, neither NOS principle 1 nor NOS principle 2 were significant predictors.

Within the first count regression model, performance on the SEB survey was examined to determine if it significantly predicted the total number of actions. SEB was entered into the model in two different ways to see which one was more appropriate for the final model. First, each of the four SEB domains was entered as four separate variables; second, the SEB total average score was entered as one variable. Keeping everything else the same, both models were fitted and the count regression model with the total average SEB (model deviance: 98.5481, AICC: 268.6583), fitted better than the same model with four separate variables for SEB survey performance (model deviance: 97.5033, AICC: 289.4212). The same relationship existed between the models when predicting the investigative number of actions and information seeking number of actions. Thus, for all three of the Poisson count regression models, the total average

Table 4 Parameter estimates of Poisson count regression for number of actions

Analysis of maximum likelihood parameter estimates

Parameter		DF	Estimate	Standard error	Wald 95% confidence limits		Wald chi-square	Pr > ChiSq
Intercept		1	− 0.3844	0.6663	− 1.6903	0.9215	0.33	0.5640
Gender	1	1	− 0.3587	0.1129	− 0.5801	− 0.1374	10.09	0.0015
Gender	2	0	0.0000	0.0000	0.0000	0.0000	.	.
Race	1	1	− 0.4641	0.1528	− 0.7637	− 0.1646	9.22	0.0024
Race	3	1	0.0489	0.1223	− 0.1908	0.2885	0.16	0.6894
Race	2	0	0.0000	0.0000	0.0000	0.0000	.	.
Expertise	1	1	1.0699	0.1494	0.7770	1.3628	51.25	<.0001
Expertise	0	0	0.0000	0.0000	0.0000	0.0000	.	.
SEB_TOTAL_AVG		1	0.7807	0.1490	0.4886	1.0728	27.44	<.0001
NOS_Principle1	2	1	0.0650	0.2166	− 0.3596	0.4895	0.09	0.7642
NOS_Principle1	3	1	0.3269	0.1920	− 0.0494	0.7032	2.90	0.0886
NOS_Principle1	4	0	0.0000	0.0000	0.0000	0.0000	.	.
NOS_Principle2	2	1	0.2348	0.2066	− 0.1702	0.6397	1.29	0.2559
NOS_Principle2	3	1	0.4116	0.1636	0.0909	0.7323	6.33	0.0119
NOS_Principle2	4	0	0.0000	0.0000	0.0000	0.0000	.	.
ExpertVerbScore		1	− 0.1090	0.0310	− 0.1698	− 0.0483	12.36	0.0004
Scale		0	1.0000	0.0000	1.0000	1.0000		

SEB was used as a predictor. Total number of actions and information seeking number of actions were significantly predicted by SEB survey performance ($\chi^2 = 27.87$, $p < .0001$ and $\chi^2 = 42.66$, $p < .0001$), respectively. However, the total number of investigative actions was not significantly predicted by SEB performance. Finally, the relationship between expert verb score and actions performed in SCI within the three Poisson count regression models discussed above was examined. Expert verb score significantly predicted the total number of actions ($\chi^2 = 12.36$, $p = .0004$) and number of investigative actions ($\chi^2 = 14.03$, $p = .0002$), but it did not contribute significantly in predicting the number of information seeking actions.

Discussion

In this study, a baseline for expert practices in a simulated authentic science inquiry environment was established. These practices may be reflective of the knowledge and beliefs about the nature of science inquiry and how inquiry generates new knowledge, or EASI. Participant practices in the simulated authentic inquiry experience offered by SCI simulations, including actions performed, type of investigation, and language used during the conclusion phases were tied to existing metrics of nature of science understanding and scientific epistemic beliefs to generate preliminary models of what student practices in authentic inquiry can reveal about underlying epistemological beliefs. Given concerns with existing metrics of NOS or science epistemology (Sandoval 2005; Sandoval et al. 2016; Sandoval and

Redman 2015), the importance of this foundational knowledge for attaining science literacy (Renken et al. 2016; Schwan et al. 2014), and the potential power of using simulations for high-throughput, real-time assessment of difficult to measure constructs, this study raised several interesting questions and prompts future research in using simulation based assessment for understanding difficult to measure constructs.

To define expert-like EASI, both experts and novices completed pre-test metrics to assess their baseline NOS understanding and science epistemic beliefs and a SCI simulation. Experts had a more sophisticated understanding of two NOS principles assessed: that science knowledge is tenuous and the lack of a universal scientific method. Experts also scored marginally higher on the SEB survey, particularly in the source and certainty domains, although this difference was not statistically significant. These results, in addition to the recruitment criteria for the experts and novices, underscores that experts not only have more experience with authentic science inquiry, but that their underlying understanding of science was also more sophisticated than novices. Since there could possibly be some concern about being able to "do" science without understanding the whole process, this result suggests that experts had both a better understanding of the process of science inquiry and experience performing inquiry. This study provides preliminary evidence of a relationship between what participants know and believe about science and their inquiry practices. Although experts were defined by their

measurable experience with authentic science practices, namely their years of experience, publishing novel primary peer-review research, and passing their comprehensive exams, it may be possible that other aspects of expertise could have influenced these results. For example, none of the experts had expertise in the specific content area of the simulation. How would practices change if an individual had content expertise, but may or may not have expertise in authentic science practices? Could an expert be more likely to use certain heuristics if they are a content expert and therefore have their investigations appear simpler? Might an expert express more naïve views within the context of the simulation even though they would express complex views in another situation (Sandoval and Redman 2015)? Perhaps one interpretation of the expert trajectories observed is that when unfamiliar with the content area experts chose a complicated and sophisticated trajectory during authentic inquiry because they could not rely on previous experience. Future work will examine expert practices both in and outside their content area of expertise.

Experts also performed more complex investigations that were aimed at uncovering a mechanistic cause and effect relationships, performed more actions total, particularly information seeking actions, and overall had inquiry profiles that alternated between periods of testing and information seeking. Notably, we observed a linear relationship between overall complexity and total actions; the increase in actions was not due to random activity during the simulation, but was likely intentional in nature. Future work with a higher sample size, and increased power, will allow fitting more advanced predictive logistic models to probe for other predictors that may contribute towards predicting the level of expertise of the participants.

The case study analysis further demonstrated that more-expert like investigations were correlated with planning time and information seeking decisions, consistent with novice/expert studies in the engineering education literature (Atman et al. 2007). The difference in information seeking decisions, both between experts and novices and between the complex and simple investigations raises some interesting questions about EASI. Why did some novices, such as the simple novice showcased here, choose to ignore the availability of information and instead say, "I do not know?" This could be explained by self-efficacy, which among high school students is related to students' scientific epistemic beliefs (Tsai et al. 2011). It also may be possible that students who do not have high self-efficacy towards their ability to do science may be unlikely to fully engage with what they do not understand, and therefore perform simple investigations. Preliminary work indicated that practices within SCI correlated with affective factors such as

self-efficacy, metacognition, or a sense of having a science identity (Peffer et al. 2018). This result could also indicate that the novices in question have only been exposed to the canned recipe-like simple inquiry that tends to predominate in K-16 classrooms (Chinn and Malhotra 2002) and without an obvious answer the default is to say "I don't know." Since simple inquiry environments are built on following a set of instructions, not engaging in independent exploration, it may be that novices have not learned about the importance of synthesizing outside information as part of their autonomous investigation, which is considered part of understanding of the nature of science inquiry (Lederman et al. 2014).

The complex expert highlighted here spent considerable time during her think-aloud to discuss the source of articles that she used during her investigation. She also reviewed the primary literature as part of her investigation, which was observed among some of the experts, but not among any novices. Source of knowledge is a dimension of scientific epistemic beliefs (Conley et al. 2004), and how online sources are evaluated is thought to represents students' epistemic thinking (Barzilai and Zohar 2012). Therefore, how students evaluate and choose to incorporate evidence into their investigations may be an important epistemologically salient episode to target for both assessment purposes and classroom instruction.

Both score on the second NOS principle (tenuous nature of science knowledge) and SEB total score predicted total actions. Since the two domains of the SEB survey that were most different between experts and novices were source and certainty, this suggested a connection between how students conceptualize the nature of science knowledge and the depth of their investigations. Although neither alone predicted investigative or information seeking actions, the presence of more actions and the correlation with overall complexity of investigation may indicate that students who have more sophisticated understanding of the nature of science knowledge engage in more sophisticated investigations aimed at revealing cause and effect relationships, not simply arriving at a single answer. However, these data are limited by a smaller sample size, and although these prospective correlations are interesting, additional work is necessary to fully understand the connections between baseline understanding of NOS and epistemological beliefs about science and how knowledge or beliefs plays out during authentic science inquiry.

Limitations
A possible confound for this study may be the interest in the topic at hand. A more interested student may have been more willing to seek information and pursue many tests to reach their conclusion. For example, one

research participant stated in his post-simulation interview that he "didn't want to commit the time to really figure it out and would stay and work more if he was earning more class credit." He also commented that he would "get an A" for doing it, regardless of how well he tried. In contrast to this participant, Sally stated multiple times in her think-aloud transcript how "interested" she was in the subject. It is also worth noting that both of the described instances were coded as simple novices. Related to interest, another potential confound is the perception of time. Beth specifically stated in her post simulation interview that "she wasn't in a hurry, had nowhere else to be."

Another limitation was with the metrics used. In addition to reliability and validity concerns (Deng et al. 2011; Sandoval and Redman 2015), we also observed very high scores on the SEB and had low interrater reliability on the NOS metric. However, despite concerns, the hypothesis that experts would perform better than novices was statistically supported. Furthermore, in the modeling analysis, both metrics were consistent with one another and predicted total number of actions performed. Investigative and information seeking actions performed were not related to complexity of investigation, likely due to small sample size and limited power. Although there was high reliability for our pre-test metric, small sample size prevented an accurate calculation of the validity. However, preliminary quantitative analyses combined with the qualitative analysis yielded provocative results about differences in authentic science practices between experts and novices. Further studies with additional participants are warranted.

Implications and future directions

The development of sophisticated epistemological beliefs about science is an essential goal of science education and overall science literacy. However, it is difficult to assess and consequently measure epistemological beliefs. Although existing metrics provide a snapshot of students' epistemological beliefs and/or NOS understanding, there are many concerns about their reliability and validity (Deng et al. 2011; Sandoval and Redman 2015).

In addition the lengthy nature of the VNOS or VASI metrics can preclude their use in a classroom setting or with large numbers of participants. Assessing what students do in authentic science inquiry as a proxy for what students know and believe about science may provide a potential solution. In the present study, experts and novices exhibited distinct inquiry trajectories that were correlated with their scores on extant metrics of NOS and epistemological beliefs about science. Although our work had a small overall sample size, the data supported the hypothesis that practices reflect epistemological beliefs. Using a computer-based assessment and learning

analytics techniques, such as automated language analysis (Peffer and Kyle 2017), could allow for high-throughput measurement and analysis of student practices.

Improved methods of assessing students' epistemological beliefs about science may provide new pedagogical avenues to both understand and address the epistemic processes of inquiry. Certain practices in authentic inquiry such as use of tentative or hedging language (Peffer and Kyle 2017), persistent seeking of information outside the simulation, and use of complex inquiry strategies were reflective of expertise and correlated with performance on pre-test assessment of NOS understanding. Future studies expanding this work, using larger sample sizes and consequently more powerful statistical models, will provide additional details about epistemologically salient episodes in authentic science inquiry that could be pedagogically targeted. For example, the effectiveness of teaching interventions designed to increase not only student understanding of NOS, but their application, could be tested. This may be particularly useful with preservice teachers who can accurately describe NOS items in the classroom, but fail to transfer that knowledge into their classrooms (Bartos and Lederman 2014).

Novice practices existed on a continuum from less to more sophisticated (Fig. 3). This diversity highlights the variety of epistemic perspectives in a given population and provides the possibility for personalizing learning in the classroom. For example, are better pedagogical outcomes observed if less expert-like novices are paired with more expert-like novices? Do less expert-like novices respond to pedagogical interventions differently than more expert-like novices? If students who are intermediary between the novices and experts participated in this study, such as advanced undergraduate biology majors, would we observe a hybrid profile between experts and novices? Tracing the development of sophisticated epistemological beliefs overtime could help indicate which existing pedagogical interventions are most effective at promoting the development of not only content and practice knowledge, but the epistemic underpinnings. Also, the identification of what "expert" means in the context of simulated authentic inquiry may reveal new pedagogical targets for promoting the development of EASI. New pedagogical interventions could also be developed by understanding what is happening over the evolution of a student's development from a novice to an expert. Furthermore, since not all novices will in fact become experts, but still require an expert-like understanding of how science works to be a productive member of society, these prospective pedagogical interventions could lead to improved overall science literacy.

Conclusions

Despite its small sample size, this study represents the first iteration of a larger study that may change the way researchers and instructors assess the underlying philosophical foundations students have about the relationship between science inquiry and generation of new science knowledge. The qualitative results described here provide important grounding for future work developing a practices-based assessment of EASI. Performance on a pre-test metric and overall expertise predicted actions performed during inquiry and were able to identify some potentially epistemologically salient episodes for future examination with a larger sample size. This work highlights the potential of using high-throughput, real-time assessment of simulated authentic science practices as a less-constrained way of examining constructs that are traditionally difficult to assess. Examining what students do during inquiry, rather than what they say about inquiry on a standard measure, removes the philosophical assumptions that come with traditional assessments. Consequently, examining EASI in SCI may lead to new areas of pedagogical focus and techniques that improve student EASI and overall science literacy. For example, how can the inquiry profiles generated by students be used to personalize instruction? The potential power of using simulations as an assessment technique to examine multiple simultaneous users in real time is exciting, as are the implications for how simulations can be leveraged to improve science literacy.

Endnotes

[1] A necropsy is an autopsy performed on an animal.

[2] We note here that climate change is the more accurate description of what Lisa means by global warming. At another point in her think-aloud transcript, Lisa acknowledges her error.

[3] PubMed is a database of biology and medical primary literature.

Additional file

Additional file 1: Table S1. Parameter Estimates of Poisson Count Regression for Number of Investigative Actions. (PDF 275 kb)

Abbreviations

AIR: Aims and Values, Epistemic Ideals, Reliable Processes; CURE: Course-Based Undergraduate Research Experience; EASI: Epistemology in Authentic Science Inquiry; NGSS: Next Generation Science Standards; NOS: Nature of Science; SCI: Science Classroom Inquiry; SEB: Scientific epistemic belief; VASI: Views About Science Inquiry; VNOS: Views of Nature of Science

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]College of Natural and Health Sciences, University of Northern Colorado, Ross Hall 1556, Campus Box 92, Greeley, CO 80639, USA. [2]Volgenau School of Engineering, George Mason University, 4400 University Drive, MS 4A7, Fairfax, VA 22030, USA.

References

Abd-El-Khalick, F. (2012). Examining the sources for our understandings about science: Enduring conflations and critical issues in research on nature of science in science education. *International Journal of Science Education, 34*(3), 353–374.

Akerson, V. L., Cullen, T. A., & Hanson, D. L. (2010). Experienced teachers' strategies for assessing nature of science conceptions in the elementary classroom. *Journal of Science Teacher Education, 21*(6), 723–745. https://doi.org/10.1007/s10972-010-9208-x.

Atman, C., Adams, R., Cardella, M., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: a comparison of students and expert practitioners. *Journal of Engineering Education, 96*(4), 359–379. https://doi.org/10.1002/j.2168-9830.2007.tb00945.x.

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., et al. (2014). Assessment of course-based undergraduate research experiences: a meeting report. *CBE Life Sciences Education, 13*(1), 29–40. https://doi.org/10.1187/cbe.14-01-0004.

Bartos, S. A., & Lederman, N. G. (2014). Teachers' knowledge structures for nature of science and scientific inquiry: Conceptions and classroom practice. *Journal of Research in Science Teaching, 51*(9), 1150–1184. https://doi.org/10.1002/tea.21168.

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: evaluating and integrating online sources. *Cognition and Instruction, 30*(1), 39–85. https://doi.org/10.1080/07370008.2011.636495.

Campbell, T., Oh, P. S., & Neilson, D. (2012). Discursive modes and their pedagogical functions in model-based inquiry (MBI) classrooms. *International Journal of Science Education, 34*(15), 2393–2419. https://doi.org/10.1080/09500693.2012.704552.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175–218. https://doi.org/10.1002/sce.10001.

Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. In D. Rapp & J. Braasch (Eds.), *Processing inaccurate information*. Cambridge, MA: MIT Press.

Conley, A. M., Pintrich, P. R., Vekiri, I., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary

Educational Psychology, 29(2), 186–204. https://doi.org/10.1016/j.cedpsych.2004.01.004.

Corwin, L., Graham, M., & Dolan, E. (2015). Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. *CBE Life Sciences Education, 14*(1), es1. https://doi.org/10.1187/cbe.14-10-0167.

Creswell, J. W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks: SAGE Publications.

Creswell, J. W., & Plano, C. V. L. (2011). *Designing and conducting mixed methods research*. Los Angeles: SAGE Publications.

Deng, F., Chen, D., Tsai, C., & Chai, C. S. (2011). Students' views of the nature of science: a critical review of research. *Science Education, 95*(6), 961–999. https://doi.org/10.1002/sce.20460.

Elby, A., Macrander, C., & Hammer, D. (2016). Epistemic cognition in science. In I. Bråten, J. Greene, & W. Sandoval (Eds.), *Handbook of Epistemic Cognition* (pp. 113–127). New York: Routledge.

Ford, M. (2015). Educational implications of choosing "Practice" to describe science in the Next Generation Science Standards. *Science Education, 99*(6), 1041–1048.

Greene, J. A., Sandoval, W. A., & Bråten, I. (2016). *Handbook of epistemic cognition*. London: Routledge Ltd. - M.U.A. https://doi.org/10.4324/9781315795225.

Hanauer, D. I., Hatfull, G. F., & Jacobs-Sera, D. (2009). *Active assessment: assessing scientific inquiry* (1st ed.). New York; London: Springer. https://doi.org/10.1007/978-0-387-89,649-6.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88–140. https://doi.org/10.2307/1170620.

Hu, D., & Rebello, N. (2014). Shifting college students' epistemological framing using hypothetical debate problems. *Physical Review Special Topics - Physics Education Research, 10*(1), 010117. https://doi.org/10.1103/PhysRevSTPER.10.010117.

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *Journal of Learning Analytics, 1*(2), 23–47. https://doi.org/10.18608/jla.2014.12.3.

Koerber, S., Osterhaus, C., & Sodian, B. (2015). Testing primary-school children's understanding of the nature of science. *British Journal of Developmental Psychology, 33*(1), 57–72. https://doi.org/10.1111/bjdp.12067.

Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire. *Journal of Research in Science Teaching, 51*(1), 65–83. https://doi.org/10.1002/tea.21125.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. E. S. (2002). Views of nature of science questionnaire: toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6), 497–521. https://doi.org/10.1002/tea.10034.

Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction, 21*(1), 137–151. https://doi.org/10.1016/j.learninstruc.2010.01.001.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, D.C.: The National Academies Press.

Osborne, J. (2014a). Scientific practices and inquiry in the science classroom. In N. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education*. Abingdon: Routledge.

Osborne, J. (2014b). Teaching scientific practices: meeting the challenge of change. *Journal of Science Teacher Education, 25*(2), 177–196. https://doi.org/10.1007/s10972-014-9384-1.

Peffer, M., Royse, E., & Abelein, H. (2018). Influence of affective factors on practices in simulated authentic science inquiry. In *Rethinking Learning in Digital Age: Making the Learning Sciences Count. Conference Proceedings of the International Conference of the Learning Sciences* (Vol. 3).

Peffer, M. E., Beckler, M. L., Schunn, C., Renken, M., & Revak, A. (2015). Science classroom inquiry (SCI) simulations: a novel method to scaffold science learning. *PLoS One, 10*(3), e0120638.

Peffer, M. E., & Kyle, K. (2017). Assessment of language in authentic science inquiry reveals putative differences in epistemology. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 138–142). New York: ACM.

Renken, M., Peffer, M., Otrel-Cass, K., Girault, I., & Chiocarriello, A. (2016). *Simulations as scaffolds in science education*. Springer International Publishing: Springer.

Rowland, S., Pedwell, R., Lawrie, G., Lovie-Toon, J., & Hung, Y. (2016). Do we need to design course-based undergraduate research experiences for authenticity? *Cell Biology Education, 15*(4), ar79. https://doi.org/10.1187/cbe.16-02-0102.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education, 89*(4), 634–656. https://doi.org/10.1002/sce.20065.

Sandoval, W. A., Greene, J. A., & Bråten, I. (2016). Understanding and promoting thinking about knowledge: origins, issues, and future directions of research on epistemic cognition. *Review of Research in Education, 40*(1), 457–496. https://doi.org/10.3102/0091732X16669319.

Sandoval, W. A., & Redman, E. H. (2015). The contextual nature of scientists' views of theories, experimentation, and their coordination. *Science & Education, 24*(9), 1079–1102. https://doi.org/10.1007/s11191-015-9787-1.

SAS Institute Inc. (2014). *SAS/ETS®13.2 User's Guide*. Cary: SAS Institute Inc.

Schizas, D., Psillos, D., & Stamou, G. (2016). Nature of science or nature of the sciences? *Science Education, 100*(4), 706–733. https://doi.org/10.1002/sce.21216.

Schraw, G. (2013). Conceptual integration and measurement of epistemological and ontological beliefs in educational research. *ISRN Education, 2013*, 1–19.

Schwan, S., Grajal, A., & Lewalter, D. (2014). Understanding and engagement in places of science experience: Science museums, science centers, zoos, and aquariums. *Educational Psychologist, 49*(2), 70–85. https://doi.org/10.1080/00461520.2014.917588.

Schwartz, R., & Lederman, N. (2008). What scientists say: scientists' views of nature of science and relation to science context. *International Journal of Science Education, 30*(6), 727–771. https://doi.org/10.1080/09500690701225801.

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: a meta-analysis. *Developmental Review, 39*, 37–63.

Someren, M. V., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: a practical approach to modelling cognitive processes*. Academic Press.

Stathopoulou, C., & Vosniadou, S. (2007). Exploring the relationship between physics-related epistemological beliefs and physics understanding. *Contemporary Educational Psychology, 32*, 255–281.

Tsai, C., Jessie Ho, H. N., Liang, J., & Lin, H. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction, 21*(6), 757–769. https://doi.org/10.1016/j.learninstruc.2011.05.002.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: model-based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*(5), 941–967. https://doi.org/10.1002/sce.20259.

Worsley, M., & Blikstein, P. (2014). Analyzing engineering design through the lens of computation. *Journal of Learning Analytics, 1*(2), 151–186. https://doi.org/10.18608/jla.2014.12.8.

Yang, F., Huang, R., & Tsai, I. (2016). The effects of epistemic beliefs in science and gender difference on university students' science-text reading: An eye-tracking study. *International Journal of Science and Mathematics Education, 14*(3), 473–498. https://doi.org/10.1007/s10763-014-9578-1.

Zeineddin, A., & Abd-El-Khalick, F. (2010). Scientific reasoning and epistemological commitments: Coordination of theory and evidence among college science students. *Journal of Research in Science Teaching, 47*(9), 1064–1093.