

Learning Analytics to Assess Beliefs about Science: Evolution of Expertise as Seen through Biological Inquiry

Melanie E. Peffer,^{†§*} Niloofar Ramezani,[‡] David Quigley,[§] Emily Royse,[†] and Chloe Bruce[†]

[†]Institute of Cognitive Science, Molecular, Cellular, and Developmental Biology, [‡]Institute of Cognitive Science, and [§]Applied Mathematics, University of Colorado Boulder, Boulder, CO 80309; [†]School of Biological Sciences, University of Northern Colorado, Greeley, CO 80639; [‡]Department of Statistics, Volgenau School of Engineering, George Mason University, Fairfax, VA 22030

ABSTRACT

Epistemological beliefs about science (EBAS) or beliefs about the nature of science knowledge, and how that knowledge is generated during inquiry, are an essential yet difficult to assess component of science literacy. Leveraging learning analytics to capture and analyze student practices in simulated or game-based authentic science activities is a potential avenue for assessing EBAS. Our previous work characterized inquiry practices of experts and novices engaged in simulated authentic science inquiry and suggested that practices may reflect EBAS. Here, we extend our prior qualitative work to quantitatively examine differences in practices and EBAS between non-science majors, biology majors, and biology graduates. We observed that inquiry practices of non-science majors and biology graduates were similar to the novice and expert practices, respectively, in our prior work. However, biology majors sometimes appeared to act like their undergraduate peers (e.g., performing fewer planning actions) but other times were more similar to biology graduates (e.g., performing complex investigations). We noted that cognitive constructs like metacognition were also important for understanding which practices were most likely to be reflective of EBAS. This work advances how to assess EBAS using learning analytics and raises questions regarding the development of cognitive processes like EBAS among aspiring biologists.

INTRODUCTION

Science literacy, or the fundamental science knowledge needed by all members of society and the skills necessary to leverage that knowledge to make informed scientific decisions (Organisation for Economic Co-operation and Development, 2006; Crowell and Schunn, 2016), is an essential goal of science education. Within biology classrooms, students need the foundational knowledge necessary to grapple with biological issues facing today's society, ranging from genetic privacy to climate change. Underlying the attainment of science literacy are several cognitive constructs, including epistemological beliefs about science (EBAS). EBAS are beliefs that an individual possesses regarding the nature of science (NOS) knowledge and how that knowledge is generated through inquiry. These beliefs influence how students learn about science and their activities in science practices, like argumentation (Elby *et al.*, 2016) or inquiry (Peffer and Ramezani, 2019). Although development of sophisticated EBAS is acknowledged as important for attainment of science literacy, EBAS are difficult to define and consequently assess. Part of the challenge lies in that epistemological beliefs are not directly observable. Because they cannot be observed directly, indirect metrics must be used, which then raises concerns about the accuracy and validity of the measurement (Ifenthaler, 2012). Current pen-and-paper metrics of EBAS and the

Ido Davidesco, *Monitoring Editor*

Submitted Nov 19, 2019; Revised Jul 2, 2020; Accepted Jul 9, 2020

CBE Life Sci Educ September 1, 2020 19:ar47
DOI:10.1187/cbe.19-11-0247

*Address correspondence to: Melanie E. Peffer (melanie.peffer@colorado.edu).

© 2020 M. E. Peffer *et al.* CBE—Life Sciences Education © 2020 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

related construct, NOS understanding, are criticized for their lack of reliability and validity, with some researchers calling for a cessation of their use (Sandoval, 2005; Sandoval and Redman, 2015).

A solution to this problem is to allow individuals to externalize these constructs. Prior work in science education suggested viewing EBAS in the context of an authentic task, such as argumentation (Mason and Scirica, 2006; Deng *et al.*, 2011) or inquiry (Sandoval, 2005; Peffer and Ramezani, 2019). Others have posited that new technological and methodological advances, such as those provided by learning analytics, can be harnessed to assess these important but difficult to measure constructs in a faster, more reliable manner (Ifenthaler, 2012; Knight *et al.*, 2014). Big data tools and methods are revolutionizing research in a wide range of disciplines, including both biology with next-generation sequencing and education via learning analytics. The field of learning analytics is relatively new, with the 10th anniversary of the first learning analytics conference being held in 2020. Learning analytics are defined in a variety of manners. Lockyer *et al.* (2013) define learning analytics as data about learners and/or learning environments that are studied and leveraged to improve learning and/or learning environments. Jisc, or the Joint Information Systems Committee, a nonprofit in the United Kingdom, defines learning analytics as the use of data about students and their activities to understand and improve educational processes and provide better support to learners (Jisc, 2015).

Here we used learning analytics to assess EBAS situated in authentic science inquiry. We examined the differences in practices and epistemological beliefs in a computer-based biological inquiry activity between non-science majors, biology majors, and individuals possessing at least one (if not more) biology degrees. We found that non-science majors tended to perform activities consistent with the novices in our prior qualitative work, whereas biology graduate practices were consistent with the experts (Peffer and Ramezani, 2019). However, biology majors sometimes appeared like the novices and at other times more like experts. This may suggest that there is a progression of EBAS that occurs during the development of an aspiring biologist that does not seem to be the result of enhanced biology content knowledge. The differences in practices observed do not seem to be the result of factors such as motivation to complete a science task or the ability to regulate one's learning activities in a science setting but could be the result of other cognitive constructs related to epistemology, such as metacognition. The work presented here is interdisciplinary in that it includes the perspective of researchers with firsthand understanding of the enculturation process of becoming a biologist along with methodologies and theories from both the learning analytics and learning sciences communities.

Theoretical Framework

Defining Science Knowledge and Scientific Inquiry. Within the literature, there are two theoretical frameworks for defining what constitutes science knowledge and how generation of that knowledge is unique from other domains of inquiry, such as religion or philosophy. Generally speaking, within the science education literature, it is called nature of science, or NOS, understanding, and in the psychology literature, EBAS. Lederman and colleagues describe NOS as the principles and beliefs

that undergird the practice of science and how science can be used as a way of knowing (Lederman *et al.*, 2002). The authors also stated that there are several key aspects for differentiating science from other domains of inquiry, like philosophy or religion, that require particular pedagogical attention. These include the tentative NOS knowledge and the empirical NOS. EBAS are operationalized in a variety of manners, including “scientific epistemic beliefs,” “personal epistemology,” and “epistemic cognition” (Hofer and Pintrich, 1997; Elby *et al.*, 2016). At a fundamental level, all of these different ways of operationalizing EBAS are founded on the beliefs an individual has about what science knowledge is and how we “know we know” scientific knowledge. Personal epistemology and scientific epistemic beliefs deal more with what beliefs an individual possesses, whereas epistemic cognition focuses on an individual's reasoning and consideration of knowing how we know.

Some have suggested that NOS and personal epistemology may be interchangeable, particularly within the context of inquiry (Deng *et al.*, 2011; Elby *et al.*, 2016). Our prior work describes the relationship between the two as bidirectional, with what you know about science (your NOS understanding) influencing what you believe about science (your epistemological beliefs) and vice versa (Peffer and Ramezani, 2019). Some aspects of NOS and EBAS overlap. For example, justification, an aspect of epistemic beliefs about science identified by Hofer and Pintrich (1997), or using evidence to support scientific claims, is very similar to the NOS principle of the empirical NOS knowledge. Another scientific epistemic belief, certainty, or how science knowledge changes over time, is similar to the NOS principle of the tentativeness of science knowledge. Because these two ways of describing NOS knowledge are both overlapping and distinct, we will refer to this cognitive construct as NOS/EBAS through the article.

In addition to the wide range of ways of operationalizing NOS/EBAS, there is a lack of consensus regarding what precisely to teach to students. For example, do we teach a universal NOS understanding (Abd-El-Khalick, 2012; Schizas *et al.*, 2016)? Or does it need to vary depending on the discipline (Lederman *et al.*, 2002; McComas, 2015)? Or is it best to focus on what aspects of NOS align with standards documents, such as the Next Generation Science Standards (McComas, 2015)? Furthermore, practicing scientists are not consistent in how they operationalize NOS (Schwartz and Lederman, 2008; Sandoval and Redman, 2015) and may possess naïve EBAS (Wong and Hodson, 2009, 2010). What makes an epistemological belief “sophisticated” is a matter of debate as well.

Given the difficulties with defining NOS/EBAS, it is not surprising that there are multiple reliability, validity, and practical concerns with existing pen-and-paper metrics. In fact, some have said these metrics should no longer be used (Sandoval *et al.*, 2016). Convergent metrics such as the Scientific Epistemic Beliefs survey (Conley *et al.*, 2014) or the Views of Science and Technology Survey (VOSTS; Aikenhead and Ryan, 1992) are easy to administer, but raise questions about forcing student responses into a “box” that does not represent the array of possible answers and whether the students are interpreting the questions as the metric authors intended (Sandoval, 2005; Sandoval and Redman, 2015). Open-ended metrics such as the Views of the Nature of Science (VNOS; Lederman *et al.*, 2002) or Views About Science Inquiry (VASI; Lederman *et al.*, 2014)

allow for a wider variety of answers, but are lengthy for study participants to complete, making survey fatigue a concern. Later uses of the VOSTS include a mixture of convergent and open-ended responses to balance feasibility of use with allowing for a wider variety of responses (Dogana and Abd-El-Khalick, 2008).

Assessment of NOS/EBAS in Authentic Contexts. An emerging solution to these assessment challenges is to examine student science practices in real time and authentic disciplinary contexts, such as inquiry or argumentation. In the context of argumentation, one study with middle school students found that quality of student arguments correlated with the sophistication of their epistemological beliefs (Mason and Scirica, 2006). Deng and colleagues (2011) found that NOS understanding can be assessed based on how well students argue scientific claims. For scientific inquiry, Sandoval (2005) argues that understanding the relationship between epistemological beliefs and inquiry practices is essential for understanding how students make sense of science.

Our prior work has examined the relationship between epistemological beliefs and inquiry practices within the simulated authentic science inquiry tool, Science Classroom Inquiry (SCI). SCI is a Web application that gives students a scaffolded authentic science inquiry experience within the confines of a typical classroom setting (Peffer *et al.*, 2015). The authenticity of the SCI experience is derived from its ability to model the thought processes necessary for performing an authentic science investigation (Peffer and Ramezani, 2019). Students are given complete autonomy to complete the simulation however they wish, including generation of various testing strategies and the option to revise their hypotheses (Peffer *et al.*, 2015). Using educational technologies like simulations not only can be leveraged to give students an authentic science inquiry experience free from many of the resource constraints in typical classrooms (Peffer *et al.*, 2015), but also provides a valuable source of clickstream and language data. These data can be used for assessment of difficult to measure or latent constructs, such as metacognition and EBAS (Ifenthaler, 2012; Knight *et al.*, 2014). So-called stealth assessments are an evidence-based method of incorporating assessment directly into a learning environment such as a game or simulation (Shute and Kim, 2014).

In our prior work, we noted that middle and high school students have a wide variety of strategies for completing SCI simulations, which we hypothesized could be reflective of differences in underlying EBAS (Peffer and Renken, 2015). To identify epistemologically relevant episodes in SCI, we conducted a mixed-methods analysis with experts and novices, wherein experts and novices were defined by prior experience with authentic science practices (Peffer and Ramezani, 2019). In this case, experts were all individuals who had published a first-author peer-reviewed journal publication in the natural sciences. Novices were undergraduate non-science major students with little to no experience with authentic science practices. We observed that novices and experts had distinct inquiry practices and that performance on existing metrics of NOS/EBAS was predictive of their inquiry practices. In particular, we observed that looking for information as part of their investigation, performing an investigation aimed at revealing an underlying cause and effect relationship for the phenomenon at hand, and

using hedging or tentative language like “may” and “support” when making conclusions were key expert practices (Peffer and Kyle, 2017; Peffer and Ramezani, 2019).

Current Study

In Peffer and Ramezani (2019), we noted a wide range of practices within our novice population. In particular, we noted that novices existed on a spectrum from more to less expert-like, which suggests that diversity of inquiry practices could be reflective of differences in NOS/EBAS and could serve as potential avenues to personalize instruction. However, our prior analysis was largely qualitative due to our sample size and did not control for affective components that could influence EBAS, such as self-efficacy beliefs (Tsai *et al.*, 2011) or science identity (Peffer *et al.*, 2018). Because the experts in our prior study also had more experience within biology than the novices, none of whom were majoring in biology, it was also possible that the differences observed in the earlier study could be the result of experience with biology. Given concerns with reliability of NOS/EBAS assessment, we also wished to test whether we could replicate our prior results in a different part of the country.

In this study, we expanded our original analysis to include an optimal sample size, therefore facilitating learning analytics modalities, including generation of predictive models and machine learning. This application of learning analytics methodologies to detect practices is particularly important for creating a scalable, high-throughput assessment of practices. Using this quantitative approach, we proposed the following research questions:

Research question 1. What other aspects (affective factors, experience with biology) influence practices and/or our understanding of NOS/EBAS as seen through inquiry?

Research question 2. What new insights into practices are revealed using machine learning techniques?

Research question 3: How do our populations differ in terms of their NOS/EBAS as seen through practices, and what does this tell us about the process of becoming a biologist?

METHODS

Participants

131 individuals participated in this study, including 71 non-science majors, 46 biology majors, and 15 biology graduates. All participating students were enrolled at the same midsized public research institution located in a small city in the Rocky Mountain region of the United States. Non-students, namely the postdoctoral associates included in our biology graduates sample, were recruited from several different research-intensive institutions. The non-science majors were predominantly female (80.3%), and the two predominant ethnic groups were white/European American (57.7%) and Hispanic/Latin American (19.7%) and the remainder was a mix of Black/African American, Asian/Asian American, and multiracial; 4.2% of students declined to identify. Non-science majors were 53.5% freshmen, 26.8% sophomores, 5.6% juniors, and 14.1% seniors. Biology majors were 58.7% female, and the two predominant ethnic groups were white/European American (65.2%) and Hispanic/Latin American (41.3%), with the rest a

mix of Black/African American, Asian/Asian American, Native American, and multiracial students. The majority of majors were in their senior or junior year (71.7% and 17.4%, respectively), with 8.7% sophomores and 2.2% freshmen. All study procedures were performed in accordance with Institutional Review Board protocol 1106538.

Among the biology graduates, there were five master's students, five doctoral students, and five postdoctoral associates. We detected no statistically significant difference between any of our biology graduate populations on any of the pretest assessments or inquiry practices analyzed, and therefore grouped them together into one category of biology graduates for analysis. The biology graduates were predominantly female (80%), and the two predominant ethnic groups were white/European American (46.7%) and Asian/Asian American (20%), with the remainder of participants being Hispanic/Latin American, Middle Eastern, or multiethnic.

The majority of non-science majors were recruited through M.P.'s non-science majors biology course, which included instruction on NOS, and received extra credit for their participation in this study. The remaining non-science majors were recruited via word of mouth and were entered in a raffle to receive a gift card. Non-science majors were defined as students from programs such as music, education, business, criminal justice, and psychology who also needed to take a certain number of science credit hours. Biology majors and some of the biology graduates were recruited from upper-division biology courses and completed study requirements as part of course activities. Biology majors were defined as students who had officially declared biology as their major, whereas biology graduates already had a degree in biology and were students in a biology graduate program. Postdoctoral associates participating in the study had a doctoral degree at the time of the study. Other biology graduates were recruited via word of mouth and were compensated for their time with a \$20 gift card.

Data Collection

All data were collected during a single meeting that lasted approximately 1–2 hours. Although the majority of participants completed activities in either a classroom or laboratory setting on campus under the supervision of a member of the research team, a few of the biology graduates participated virtually via Web conferencing software. Some of the biology majors completed the pretest before attending class, and the simulation was performed under the supervision of a member of the research team. First, participants completed a pretest that included both the motivation and learning strategies items from the Motivated Strategies and Learning Questionnaire (MSLQ; Pintrich, 1991), and also an assessment of science identity, or how much they identified as a “science person” (Hazari *et al.*, 2010; Cribbs *et al.*, 2015; Godwin *et al.*, 2016). Both of these surveys were Likert-scale based. MSLQ items were slightly rewritten to be specific to science classes, rather than course work in general. The MSLQ included 81 items divided into 15 subscales about students' motivation and use of learning strategies. Motivation items included both intrinsic and extrinsic motivation (known respectively as intrinsic and extrinsic goal orientation), student's evaluation of interest in and/or utility of a task (task value), students' beliefs that their efforts will result in a positive outcome (control of learning beliefs), self-assess-

ment of competency toward performing a science task (self-efficacy), and concern over performance (test anxiety). Learning strategies items included skills such as practicing to learn information (rehearsal), building connections between new and prior knowledge (elaboration), logically structuring knowledge (organization), applying prior knowledge in new situations (critical thinking), as well as participants' awareness of their own thinking (metacognition), how to create an environment conducive for studying (time and study environment), ability to control their efforts toward attaining a goal (effort regulation), ability to seek help from peers (peer learning) or others, such as instructors (help seeking). The science identity metric included 12 items divided into three subscales representing contributors to science identity: recognition from others of being a “science person,” feelings of competence when learning science, and interest in science. Taken together, these factors inform one's self-assessment of seeing oneself as a science person (Carlone and Johnson, 2007).

Participants also completed two assessments of NOS/EBAS, the multiple-choice VOSTS (Aikenhead and Ryan, 1992) and an open-ended NOS assessment with modifications suggested by Dogan and Abd-El-Khalick (2008) to include the option for open-ended responses. The NOS assessment included items originally published on either the Views of the Nature of Science (VNOS; Lederman *et al.*, 2002) or VASI (Lederman *et al.*, 2014; Supplemental Table 1). We decided not to include the full versions of either the VNOS or VASI, because not all aspects were relevant to our study and survey fatigue was a concern. Items were chosen based on our prior work with this instrument and the SCI simulations (Peffer and Ramezani, 2019) and assessed on two NOS principles: principle 1, the lack of a universal scientific method; and principle 2, the tentative NOS knowledge. These aspects were chosen because they are reflective of both EBAS and NOS theory. Sophisticated scores were in line with current scholarship on EBAS and NOS. For example, acknowledging that scientific knowledge is subject to change in light of new evidence. Naïve scores were opposite of what is accepted in the literature; for example, stating that science knowledge never changes. Mixed responses reflected an understanding that was both in line with the literature and the opposite of accepted theory. Pretest items were counterbalanced, and the MSLQ included items that were reverse coded.

We opted not to include any of the VOSTS results in our analyses due to concerns about reliability and validity. In this study, individual VOSTS items' reliability measures as well as the overall Cronbach's alpha of the VOSTS instrument, which was 0.45, were poor. Our exploratory factor analysis, used to establish a preliminary construct validity, demonstrated the majority of items either cross-loaded (indicating nonspecificity of VOSTS items to assess what we wanted) or had low factor loading (indicating that the factors were not strongly reflective of the underlying construct we wanted to assess).

After completing the pretest, students were instructed to activate Windows' Steps Recorder and launch the SCI simulation. SCI captured all actions and open-ended responses to embedded questions within the simulation, and the Windows' Steps Recorder captured all information-seeking activities from outside the simulation. These two data streams were merged after data collection was completed to generate a single complete log file for each user. All participants completed “The

Invasion of the Grackles” SCI simulation. In this simulation, students were tasked with determining a cause for the range expansion of a nuisance bird species, the great-tailed grackle. Originally from South America, great-tailed grackles are moving northward. Much like other SCI simulations, the authenticity of the experience is derived from the lack of a single answer to explain this phenomenon and the complete autonomy given to the students to generate their own hypotheses, revise their hypotheses, pursue a testing strategy, and decide when to conclude. The version of SCI completed by the simulation was an upgraded version of the Web app used in Peffer and Ramezani (2019). Although the overall design and flow of the simulation was the same, the user interface was streamlined. Demographic information was collected after students had completed the simulation to avoid any potential stereotype threat.

Data Analysis

Pretest Metrics. The MSLQ was coded as described in the scoring guide (Pintrich, 1991), and an average score for each construct was calculated. For the science identity metric, participant scores for each of the three subscales (competence, interest, and recognition) were averaged, and then the mean of those scores was calculated as a “proxy variable” to overall science identity (Wang and Hazari, 2018). The proxy variable was significantly and positively correlated ($r = 0.91$, $n = 117$, $p < 0.001$) to the self-recognition item of the metric (“I see myself as a science person”), suggesting that this proxy variable is valid to use as a measure of science identity, and therefore we only used the proxy variable in our analyses.

For the open-ended NOS items, as in our prior work, two members of the research team (M.P. and E.R.) coded all open-ended responses based on two NOS principles, the lack of a universal scientific method (principle 1) and the tenuous NOS knowledge (principle 2). Coding was blinded, and overall agreement was 64% for principle 1 and 61% for principle 2. Kappa values were 0.31 and 0.38, respectively, indicating fair agreement (McHugh, 2012). Disagreements were settled through mutual discussion.

SCI Practices

Prior work has suggested that clickstream or trace data, that is, the activity records of the participants as they engage with SCI, can be mapped to theoretically relevant activities (Quigley *et al.*, 2017) or cognitive constructs (Ifenthaler, 2012) in a real-world activity, including SCI (Peffer and Ramezani, 2019; Peffer *et al.*, 2019). As in our prior work, actions within SCI were categorized as either investigative, information seeking, or planning. Investigative actions included generation of a hypothesis (H), performing a test (T), or concluding (C). Information-seeking actions (I) included any time a participant looked for information as part of the investigation, such as through the internal simulation library or external to the simulation, such as through Internet search engines. Planning actions were defined as any information seeking that occurred before the generation of the first hypothesis. We were particularly interested in planning actions, as both our prior qualitative work with SCI (Peffer and Ramezani, 2019) and expert/novice studies in engineering (Atman *et al.*, 2007) indicate that planning is an expert-like practice within an authentic activity. These actions were also chosen not only because they are important parts of science

inquiry, but because certain aspects, such as the decision to search for more information, could be reflective of underlying epistemological beliefs about the source of science knowledge (see Peffer and Ramezani, 2019).

Using the log data for each participant, we counted the number of each action type, including the number of each individual action type and total number of actions. Because the total number of actions could vary among participants, we also calculated the relative rate of each action (i.e., count of that action type from that participant divided by the total number of actions from that participant). In addition, we calculated bigrams of actions (e.g., IT would represent an individual looking for information, I, immediately before performing a test, T) and maximum repeated I or T actions in a row. This was important, because, as identified in our prior qualitative work on SCI, experts often switched between testing and looking for information and/or had long information-seeking or planning phases (Peffer and Ramezani, 2019).

As in our prior work (Peffer and Ramezani, 2019), we also assessed investigative strategy. Each log file was assessed by two independent coders as simple or complex in nature. Simple investigations were neither systematic or mechanistic and were reminiscent of simple inquiry as described by Chinn and Malhotra (2002). Simple inquiry is straightforward and generally does not include iteration or offers an explanation for the observed phenomenon that includes any information regarding an underlying mechanism. For example, one simple investigation began with the hypothesis “The reason for the great-tailed grackle’s range expansion is the changing climate in their original homes.” The user then performed two tests (examining temperature and moisture data in original vs. expanded range) before concluding that “precipitation is the main cause of the great-tailed grackle’s range expansion,” because “where they are migrating from has extreme precipitation as compared with where they are migrating to. They are migrating to drier places.”

In contrast, complex investigations were often geared toward finding an underlying mechanism, logical, and evidence based. Complex investigations moved beyond a simple linear relationship between a few data sources and the question at hand, seeking to connect information together in a way that tells an interconnected story. This can involve generating a mechanistic conclusion that presents multiple pieces of interconnected information or systematically exploring all aspects of a hypothesis. For example, one complex investigation opened with “due to human alteration of the environment, factors such as climate change have caused the great-tailed grackle to expand its range.” The user then proceeded to complete two tests (same as the simple investigation described earlier, except that the user examined nesting behavior and differences in temperature between original and expanded ranges). This user concluded that “climate change has had a significant influence on the grackles’ range expansion,” because “climate change is influenced by human alteration of the environment, including urbanization. Because the grackles tend to nest near human habitations, human expansion will also lead to grackle expansion.” The user in question performed the same number of tasks as in our simple example, but instead connected the information at hand in a logical manner to generate a stronger conclusion. For additional examples, our prior work included detailed case studies of simple/complex investigations (Peffer and Ramezani, 2019).

Coding disagreements were settled by mutual discussion, and Cronbach's alpha was 0.71, indicating acceptable agreement.

Cluster Analysis

We used *k*-means clustering, an unsupervised machine learning algorithm, to determine whether inquiry practices (i.e., H, T, I, C, described earlier) would cluster together. The goal of *k*-means clustering is to group samples with similar features for the purpose of revealing an underlying pattern. Therefore, analysis of the various investigative features could reveal distinct groups that are potentially representative of underlying EBAS/NOS. We extended our prior work with this data set (Peffer *et al.*, 2019) to include an expanded sample of undergraduate students as well as biology graduates. We also expanded our features used for classification beyond relative rate of action type to include two unit groups of actions, or bigrams. For the *k*-means clustering analysis, we broke the individual log files into relevant features (Table 8, discussed later in the article). Weka, a machine learning tool (Hall *et al.*, 2009), was then used to generate emergent clusters. Features used did not include demographics, major, or pretest performance. This allowed us to use these emergent activity-based clusters as a new basis in our models for exploring both connections to the demographic features and differences in outcome measures. The elbow method (Ketchen and Shook, 1996) was used for determining the optimal number of clusters that would best represent the data while also minimizing error. The elbow method compares how well the clusters have captured the total amount of variance within the data across different potential numbers of clusters. We observed a strong "elbow" at three clusters, and therefore based our analysis on three clusters.

Statistical Analysis

Statistical analyses were performed in SPSS v. 26, SAS v. 9.4 (SAS Institute Inc, 2014), and R (R Core Team, 2017). Figures were generated in R. The significance level of $\alpha = 0.05$ was used throughout this study. To determine differences in MSLQ and science identity between our three populations (non-science majors, biology majors, and biology graduates), we performed a one-way analysis of variance (ANOVA) to evaluate statistical differences, with a Tukey's post hoc test to detect the pairwise differences between populations. Tukey's test was developed to account for multiple comparisons and maintains the appropriate alpha level to prevent the inflation of type I error (Lee and Lee, 2018). Due to small counts for certain variables, NOS performance was compared using Fisher's exact test. One-way ANOVA tests were performed to compare differences in practices among our three populations as well, and chi-squared or Fisher's exact tests were performed, as appropriate, to compare practices to educational background. A Bonferroni correction was applied to adjust for multiple testing and avoid type I error. This adjustment is recommended (Noble, 2009) when multiple tests are performed in a study.

Predictive Analysis

We used generalized linear models, an extension of linear regression modeling used when response variables are not continuous and/or normally distributed, to build predictions of key outcomes based on learner activity within the system. Logistic regression and Poisson regression models are specific cases of

generalized linear models and were used here as predictive models. Random forests and group least absolute shrinkage and selection operator (LASSO; Kukreja *et al.*, 2006) were performed in SAS v. 9.4 to select only important variables within some of the predictive models that required a higher statistical power. Random forests and LASSO models are dimension-reduction statistical approaches appropriate to use when working with many predictor variables. We used them here, as they can also be used as variable selection methods before fitting predictive models. Next, the important variables were entered into logistic regression models in SAS v. 9.4 to build the predictive models and identify the variables that were significant contributors in the modeling of the binary and multinomial/categorical response variables. Because predicting simple or complex investigations is a dichotomous task, we used a binary logistic regression model. Similarly, for predicting cluster assignment (which was a categorical variable with three populations), we used a multinomial logistic regression model. When modeling the count response, we had sufficient power to keep all the predictors in the model. We fit a Poisson (count) regression model, in SAS v. 9.4, to predict both the total number of actions and planning actions performed.

RESULTS

Baseline Differences between Non-science Majors, Biology Majors, and Biology Graduates

Because cognitive constructs like self-efficacy can influence epistemological beliefs (Tsai *et al.*, 2011), we first compared baseline differences in motivation (including self-efficacy), learning strategies, and science identity between our three populations of interest. A one-way between-subjects ANOVA was conducted to compare the pretest performance between our three populations of interest with a Tukey's post hoc test to identify differences within samples. We found that each cognitive construct fell into one of four groups (Figure 1 and Table 1). Within group 1, there was no difference between our three populations (extrinsic goal orientation, rehearsal, time and study environment, help seeking). We also note only marginal significance for control beliefs and effort regulation. Group 2 represents cognitive constructs that are very similar among undergraduates regardless of major, but differ from those of biology graduates (Figure 1A). These include aspects that were similar among the undergraduates, but different from those of biology graduates, such as test anxiety, organization, effort regulation (marginally significant), and metacognition (marginally significant). Group 3 represents cognitive constructs that were very similar between biology majors and biology grads, but different from those of non-science majors (Figure 1B). These included intrinsic goal orientation, task value, elaboration, critical thinking, and science identity. Group 4 was a catch-all category and included self-efficacy, which was similar between biology majors and biology graduates, but lower for non-science majors. We also noted that, for peer learning, biology majors and non-science majors differed, but there was no difference between either undergraduate population or the biology graduates (Figure 1C).

We next looked at differences between each population and performance on the NOS assessment. Table 2 shows the results of the NOS assessment. Due to small counts for some observations, a chi-square test of association was no longer appropriate, and instead Fisher's exact test was used. We observed no

TABLE 1. Baseline differences on motivation, learning strategies, and science identity^a

Item	Non–science majors mean (SD)	Biology majors mean (SD)	Biology graduate students mean (SD)	<i>p</i> value	<i>F</i>
Motivation subscales					
Intrinsic goal orientation	3.94 (1.49)	5.10 (1.3)	5.87 (0.56)	0.00	<i>F</i> (2, 109) = 13.67
Extrinsic goal orientation	4.93 (1.22)	5.19 (1.47)	4.58 (1.23)	0.50	<i>F</i> (2,109) = 0.71
Task value	4.26 (1.36)	5.66 (1.44)	6.15 (0.63)	0.00	<i>F</i> (2, 109) = 16.80
Control beliefs	5.00 (1.19)	5.24 (1.36)	5.87 (0.75)	0.05	<i>F</i> (2, 109) = 3.13
Self-efficacy	4.34 (1.40)	4.90 (1.37)	5.86 (0.82)	0.00	<i>F</i> (2, 109) = 7.12
Test anxiety	4.43 (1.52)	4.6 (1.73)	3.15 (1.41)	0.04	<i>F</i> (2, 109) = 3.46
Learning strategies subscales					
Rehearsal	4.67 (1.16)	4.69 (1.05)	4.37 (1.55)	0.57	<i>F</i> (2,109) = 0.57
Elaboration	4.16 (1.31)	4.95 (1.12)	5.60 (0.82)	0.00	<i>F</i> (2, 109) = 8.58
Organization	4.4 (1.11)	4.67 (1.3)	5.71 (1.02)	0.00	<i>F</i> (2, 109) = 5.81
Critical thinking	3.31 (1.22)	4.31 (1.19)	5.05 (1.06)	0.00	<i>F</i> (2,109) = 13.05
Metacognition	4.14 (0.95)	4.47 (1.06)	4.91 (0.80)	0.04	<i>F</i> (2,109) = 3.26
Time and study environment	5.02 (0.98)	4.84 (1.05)	5.69 (0.89)	0.10	<i>F</i> (2,109) = 2.38
Effort regulation	4.91 (1.28)	4.25 (1.32)	5.87 (0.80)	0.05	<i>F</i> (2,109) = 3.20
Peer learning	3.51 (1.39)	4.11 (1.14)	4.03 (1.27)	0.01	<i>F</i> (2,109) = 4.42
Help seeking	4.01 (1.20)	3.27 (0.64)	4.62 (1.18)	0.17	<i>F</i> (2,109) = 1.79
Science identity					
Science identity	1.82 (0.96)	3.27 (0.64)	3.57 (0.31)	0.00	<i>F</i> (2, 109) = 42.90

^aBold type indicates statistically significant values.

statistical difference among various populations, although we do note our observation that biology graduates scored lower on average (indicating a more sophisticated answer) than either population of undergraduates on both NOS principles assessed (Table 2).

Differences in Practices between Non–science Majors, Biology Majors, and Biology Graduates

Our prior qualitative work indicated a diversity of practices between experts and novices (Peffer and Ramezani, 2019). In particular, we noted that novices varied in the amount of expert-like practices present, whereas experts were more consistent. Therefore, in this study, we wished to determine whether the diversity of practices was due to experience with subject matter. Figure 2 demonstrates the diversity of practices between our three populations organized in two different ways, first by educational background (Figure 2A) and second by cluster assignment, which is discussed later in this article (Figure 2B). Each row represents a single participant. Across all three populations (non–science majors, biology majors, and biology graduates), we see two subpopulations: those whose investigations contained planning phases before generation of their first hypothesis and those whose investigations did not contain planning phases (Figure 2A). Given that the average number of actions performed across all participants was nine, we also see that the majority of biology graduates and biology majors performed more than the average number of actions, whereas the non–science majors were more variable.

We next looked specifically at several features distinctly identified in experts in our previous work (Peffer and Ramezani, 2019) and putatively reflective of more sophisticated EBAS. First, we looked at total actions performed, as this was found in our prior work to be predictive of expertise (Table 3). The results of our one-way ANOVA indicated that the mean scores

among our three populations were significantly different, $F(2,128) = 9.40$, $p < 0.0001$. We also used a Tukey's post hoc test as a follow-up of the one-way ANOVA to identify within-sample pairwise differences (Figure 3). We noted that, for total actions, the undergraduate populations were not different from each other, only different from the biology graduates (Figure 3A).

To determine possible explanations, and contributing features, for why we observed this difference in total actions, we built a predictive model that identified which factors were associated with performing more or fewer actions. A Poisson (count) regression model was fit to figure out which variables were predictive of total number of actions (Table 4). Table 4 shows the analysis of effects results and the *p* values of each of the predictors. Supplemental Table 2 shows the parameter estimates, standard errors, confidence intervals, and hypothesis test results of every predictor within this Poisson regression model. We found that educational background, critical thinking, metacognition, time and study environment, effort regulation, science identity, NOS principle 2, gender, and whether or not the participant revised his or her hypothesis contributed significantly in predicting the total number of actions. In particular, we noted that higher scores on metacognition, effort regulation, and science identity were more likely to correspond to an increase in total numbers of actions. In contrast, increased critical thinking and time and study environment, which is how a student regulates the area where they study and how they study, were associated with predicting fewer actions.

Significance of educational background was due to the total number of actions performed by non–science major students compared with biology graduates but not the biology majors versus biology graduates (Supplemental Table 2). For NOS principle 2, the tentative NOS knowledge, the main difference

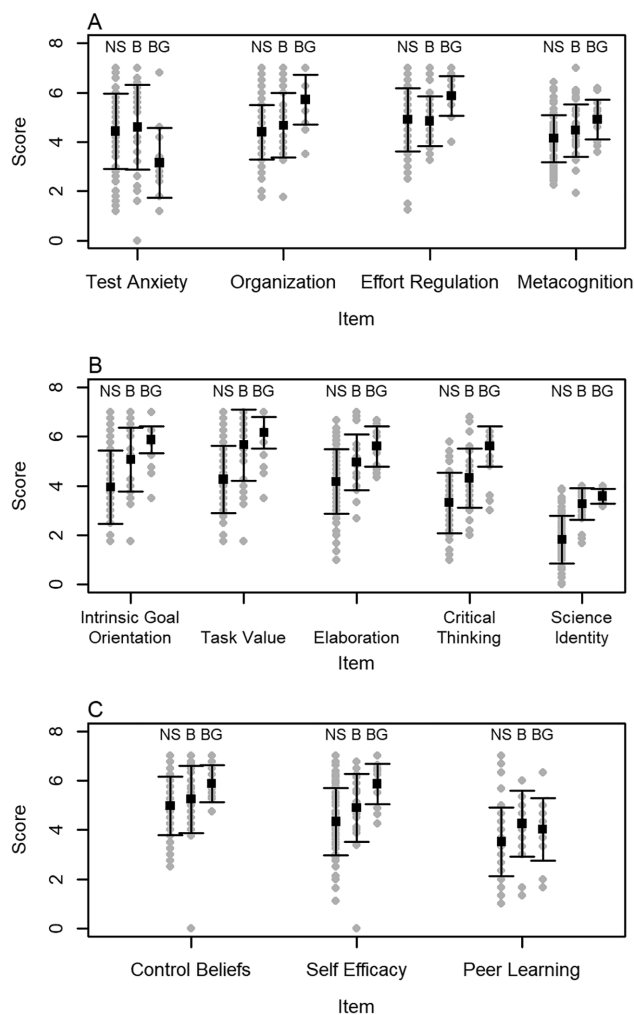


FIGURE 1. Scatter plots of participants' scores on pretest items. (A) Items for which the two undergraduate populations were similar to one another. (B) Items that were similar between biology majors and biology graduates, but different from non-science majors. (C) Constructs that were statistically different between populations, but did not fit the patterns shown in A or B. Black squares indicate means and error bars represent 1 SD. NS, non-science majors; B, biology majors; and BG, biology graduates. ANOVA tests for all items shown are statistically significant at the $\alpha = 0.05$ level, corrected for multiple comparisons.

in the total number of actions is seen between level 1 (sophisticated) and level 3 (naïve) individuals and not among level 2 (mixed) versus level 3 (naïve) participants. Said otherwise, the more sophisticated the response on NOS principle 2, the more actions a student was likely to perform.

TABLE 2. Average scores by level on NOS assessment

	Principle 1 mean (SD)	Principle 2 mean (SD)
Non-science major	2.67 (0.50)	2.07 (0.79)
Biology major	2.64 (0.59)	2.12 (0.62)
Biology graduate	2.37 (0.63)	1.69 (0.63)

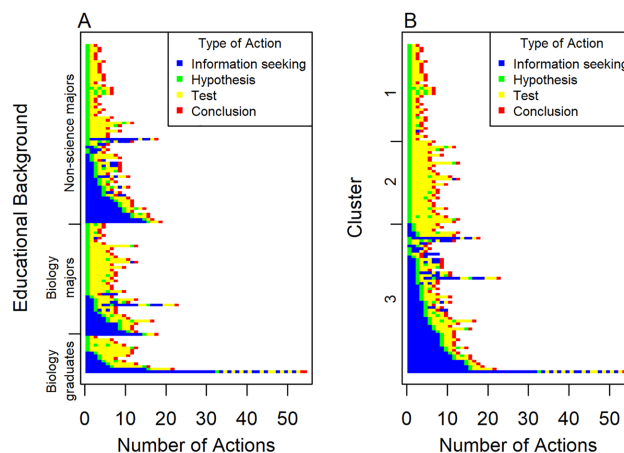


FIGURE 2. Visualization of participants' actions in SCI simulation by (A) educational background and (B) *k*-means cluster assignment. Each row represents a single participant. Within each grouping, individuals are then organized based on the total number of information-seeking actions performed.

We next looked at planning actions, or the amount of information-seeking actions that occurred before the generation of the first hypothesis, which is an expert-like process in both SCI (Peffer and Ramezani, 2019) and engineering (Atman *et al.*, 2007). Using a one-way ANOVA followed by a Tukey's post hoc test to identify within-sample differences, we noted that, for planning actions, the undergraduate populations did not differ from each other, but did differ from the biology graduates, $F(2,128) = 3.40$, $p = 0.04$ (Table 3 and Figure 3A).

To determine possible explanations for why we observed this difference in planning actions, we built another predictive model capable of identifying the significant predictors associated with performing a higher or lower number of planning actions. A Poisson (count) regression model was fit to all predictors (Table 5). Supplemental Table 3 shows additional information, such as the parameter estimates, standard errors, confidence intervals, and hypothesis test results, for this Poisson regression model. These results showed that educational background, extrinsic motivation (extrinsic goal orientation), elaboration, ability to regulate their study environment (time and study environment), peer learning, science identity, NOS principle 1, NOS principle 2, race, and gender significantly predicted the number of actions performed before generating the first hypothesis. We noted that higher scores on elaboration or increased identification as a science person, plus more sophisticated understanding of the lack of a universal scientific method (principle 1) and the tentativeness of science knowledge (principle 2) were more likely to be associated with a higher number of planning actions. In contrast, we noted that extrinsic goal regulation, time and study environment, and peer learning were most likely to be associated with decreased planning actions.

We note that the statistically significant difference we observed between educational backgrounds is due to the variation in the performance of non-science major students versus biology graduates and not the biology majors versus biology graduates (Supplemental Table 3). As seen in this table, the comparison of non-science majors with biology graduates is statistically significant ($p = 0.01$), meaning the overall

TABLE 3. Differences in actions performed among participants^a

	Planning mean (SD)	NumI mean (SD)	NumH mean (SD)	NumT mean (SD)	Total mean (SD)	RateI mean (SD)	RateH mean (SD)	RateT mean (SD)
Non-science majors	1.97 (3.57)	2.54 (3.89)	1.16 (0.61)	2.42 (1.65)	7.49 (4.08)	0.23 (0.28)	0.19 (0.10)	0.36 (0.21)
Biology majors	1.69 (3.22)	2.25 (3.69)	1.30 (0.59)	4.26 (2.44)	8.98 (4.09)	0.18 (0.25)	0.16 (0.08)	0.50 (0.23)
Biology graduates	5 (8.76)	5.71 (11.20)	1.36 (0.63)	6.14 (2.74)	14.21 (12.28)	0.24 (0.26)	0.13 (0.09)	0.51 (0.21)

^aBold type indicates statistically significant differences between groups. “Num” refers to the total number of times the action was performed, whereas “Rate” describes the relative rate at which that action type was performed across the participant’s entire investigation. I, information-seeking actions; H, hypotheses generated; and T, tests performance.

significant difference in the education background was due to the difference in performance of non-science majors versus biology graduates, rather than the biology majors versus biology graduates ($p = 0.11$). For NOS principle 1, the lack of a universal scientific method, the main difference in number of planning actions is observed between level 1 (sophisticated) and level 3 (naïve) participants and not between level 2 (mixed) and level 3 (naïve) of NOS principle 1 individuals. For NOS principle 2, the difference in number of planning actions is observed both among level 1 versus level 3 and among level 2 versus level 3. Said otherwise, the more sophisticated the understanding of the lack of a universal scientific method, the more planning actions individuals were likely to perform.

Although not observed in our prior qualitative work (Peffer and Ramezani, 2019), we also noted statistically significant differences in both the total number and relative rate of tests,

$F(2,128) = 23.80, p = 0.00$, and $F(2,128) = 6.54, p = 0.00$, respectively (Table 3). Because the relative rate and total number of tests are consistent with one another, this suggests that the variance between populations is not due to differences in total actions performed. When controlling for the total number of actions, we noticed that relative rate of testing actions was similar between biology majors and biology graduates, with non-science majors performing fewer tests (Figure 3B). We observed no statistically significant differences in either the relative rate or total number of hypotheses generated or information-seeking actions.

Because our prior work suggested that experts performed more complicated investigations with some kind of mechanistic or systematic focus (Peffer and Ramezani, 2019), we next examined differences in complexity of investigations. A chi-square test of independence was performed to examine the relationship between educational background and complexity of investigation. The relationship between the variables was significant, $\chi^2(2, n = 125) = 16.49, p < 0.0001$, and Cramer’s $V = 0.363$ indicated a large effect size (Figure 4). We noted that non-science majors performed predominantly simple investigations (59%) compared with both the biology majors (23%) and biology graduates (21%). Therefore, complexity of investigation reflected educational background.

We next performed a binary logistic regression model to examine which factors were important for predicting whether a participant would perform a simple or complex investigation. The predictive model was significant, $\chi^2 = 67.6315, p < 0.0001$, indicating that it was an efficient model in which predictors significantly contributed to forecasting whether students would perform simple or complex investigation. Additionally, this model had a good fit according to the Hosmer-Lemeshow goodness-of-fit test, $\chi^2 = 5.0610, p = 0.751$, meaning that the model fit to the data was adequate. Table 6 shows the results of this logistic regression model.

We found that participants who believed that their efforts in a science learning environment (control beliefs) would lead to a positive outcome or that their ability to self-regulate their learning (effort regulation), increased metacognitive ability and who identified as a science person were all predictive for performing a complex investigation. Conversely, we observed that, the higher the values of self-efficacy and/or help seeking are, the more likely students are to perform simple actions over complex actions. Therefore, more positive beliefs about ability to complete a science task (self-efficacy) and ability to find help in a science learning environment were predictive of performing a simple investigation. Supplemental Table 4 (results from the same logistic model presented in Table 6) shows the parameter estimates for this logistic regression model.

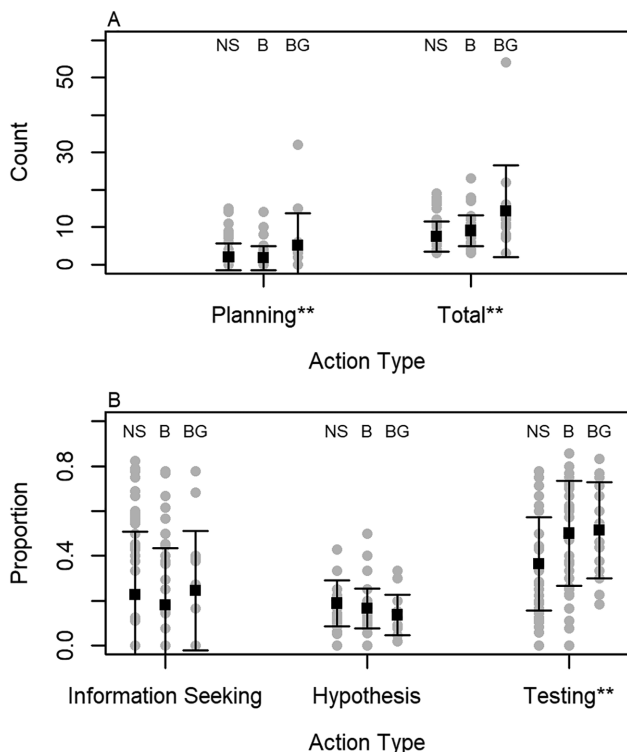


FIGURE 3. Scatter plots of (A) counts of participants’ actions in SCI simulation and (B) proportion of participants’ actions in SCI simulation. Black squares indicate means and error bars represent 1 SD. NS, non-science majors; B, biology majors; and BG, biology graduates. Double asterisk (**) indicates statistical significance at the $\alpha = 0.05$ level.

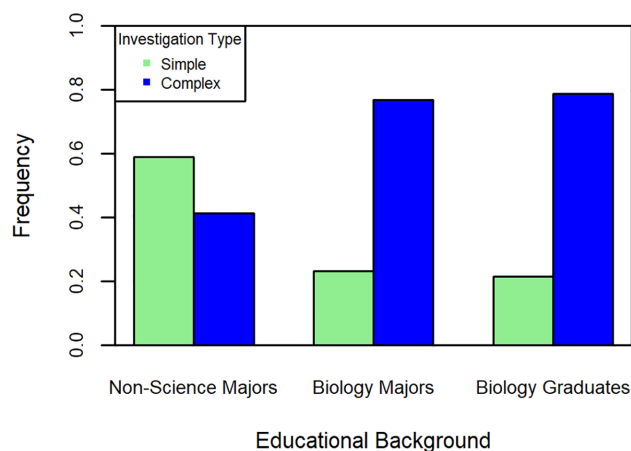
TABLE 4. Poisson regression analysis of effects for total number of actions^a

LR statistics for type 3 analysis			
Source	df	Chi-square	p-value > ChiSq
Population	2	7.82	0.0200
Intrinsic goal regulation	1	0.09	0.7666
Extrinsic goal regulation	1	3.32	0.0684
Task value	1	0.13	0.7165
Control beliefs	1	1.24	0.2662
Self-efficacy	1	0.06	0.8144
Test anxiety	1	0.05	0.8295
Rehearsal	1	0.07	0.7896
Elaboration	1	0.63	0.4259
Organization	1	0.64	0.4225
Critical thinking	1	8.36	0.0038
Metacognition	1	6.29	0.0121
Time study environment	1	28.64	0 < 0.0001
Effort regulation	1	5.31	0.0212
Peer learning	1	3.34	0.0675
Help seeking	1	0.08	0.7726
Science identity	1	6.37	0.0116
NOS principle 1	2	1.93	0.3813
NOS principle 2	2	15.43	0.0004
Race	2	4.91	0.0858
Gender	1	8.93	0.0028
Hypothesis revision	1	4.65	0.0311

^aBold font indicates statistically significant values.

Due to the high number of predictors for this logistic regression model, we ran a second model using a variable selection method (conditional backward selection). Half of the original predictors were selected for a smaller logistic model with a similar fit to the previous logistic model (Akaike information criterion = 134). LASSO confirmed this variable selection. This step was taken to provide more power to the smaller model in order to identify smaller differences. The important predictors selected through the variable selection procedure were control beliefs, self-efficacy, rehearsal, elaboration, metacognition, time and study environment, effort regulation, help seeking, science identity, NOS principle 1 (lack of a universal scientific method), and hypothesis revision.

Fitting the smaller logistic regression model to the same response as the previous model resulted in a significant model ($\chi^2 = 59.14$, $p < 0.0001$) with a good fit ($\chi^2 = 3.45$, $p = 0.903$). Table 7 shows the results of this model. The variables that significantly contributed to the prediction of whether students opt for simple or complex investigation were control beliefs, self-efficacy, rehearsal, metacognition, time and study environment, effort regulation, help seeking, science identity, and hypothesis revision. Although NOS principle 1, the lack of a universal scientific method, is not significant, we did note that category 1 (sophisticated) of NOS principle 1 compared with its category 3 (naïve) is significant in predicting complexity. The number of significant variables has increased compared with the bigger model containing 22 predictors, due to the higher statistical power of the model with only 11 predictors. The higher power of this model helped reveal more information about the significant variables in predicting the binary complexity variable.

**FIGURE 4. Distribution of investigation type (simple or complex, see Methods) by educational background.**

Supplemental Table 5 shows the parameter estimates for this logistic regression model. According to this table, the higher the control beliefs, metacognition, effort regulation, and science identity are, the more likely students are to perform a complex investigation rather than a simple investigation. On the other hand, an increase in self-efficacy, rehearsal, time and study environment, and help seeking increases the likelihood of performing simple investigations rather than complex investigations (Figure 5). Similarly, for hypothesis revision, participants who did not revise their hypotheses, compared with those

TABLE 5. Poisson regression analysis of effects for number of actions before generation of first hypothesis^a

LR statistics for type 3 analysis			
Source	df	Chi-square	p-value > ChiSq
Population	2	6.28	0.0432
Intrinsic goal orientation	1	1.74	0.1868
Extrinsic goal orientation	1	34.04	0 < 0.0001
Task value	1	0.03	0.8522
Control beliefs	1	2.51	0.1129
Self-efficacy	1	0.29	0.5929
Test anxiety	1	1.64	0.1999
Rehearsal	1	0.51	0.4757
Elaboration	1	7.67	0.0056
Organization	1	0.15	0.6988
Critical thinking	1	3.38	0.0658
Metacognition	1	0.01	0.9209
Time study environment	1	7.34	0.0067
Effort regulation	1	1.93	0.1644
Peer learning	1	13.07	0.0003
Help seeking	1	1.86	0.1729
Science identity	1	22.50	0 < 0.0001
NOS principle 1	2	8.35	0.0154
NOS principle 2	2	18.05	0.0001
Race	2	37.35	0 < 0.0001
Gender	1	30.68	0 < 0.0001
Hypothesis revision	1	1.34	0.2476

^aBold font indicates statistically significant values.

TABLE 6. Logistic regression analysis of effects predicting simple or complex investigation using all predictors^a

Type 3 analysis of effects			
Effect	df	Wald chi-square	p-value > ChiSq
Population	2	0.3141	0.8547
Intrinsic goal regulation	1	0.1843	0.6677
Extrinsic goal regulation	1	0.4870	0.4853
Task value	1	0.0761	0.7826
Control beliefs	1	5.1093	0.0238
Self-efficacy	1	5.5698	0.0183
Test anxiety	1	0.0089	0.9248
Rehearsal	1	2.1996	0.1380
Elaboration	1	1.4810	0.2236
Organization	1	0.1953	0.6585
Critical thinking	1	1.9224	0.1656
Metacognition	1	4.1078	0.0427
Time study environment	1	3.7104	0.0541
Effort regulation	1	5.8355	0.0157
Peer learning	1	1.8824	0.1701
Help seeking	1	5.6166	0.0178
Science identity	1	7.3850	0.0066
NOS principle 1	2	3.1927	0.2026
NOS principle 2	2	3.3259	0.3126
Race	2	0.8296	0.6605
Gender	1	1.2346	0.2665
Hypothesis revision	1	1.4401	0.2301

^aBold font indicates statistically significant values.

who revised their hypotheses, are more likely to perform simple investigations, as opposed to complex investigations.

Performance on NOS principle 1 was another variable that played a significant role in predicting complexity. However, the only statistically significant difference was found when comparing category 1 (sophisticated) with category 3 (naïve). The more sophisticated a participant's understanding of authentic science methods, meaning they understood that there is no universal scientific method, the more likely the participant was to perform a simple investigation. In contrast, if a participant has higher scores on NOS principle 2, meaning the more sophisti-

TABLE 7. Logistic regression analysis of effects predicting simple or complex investigation using 11 predictors^a

Type 3 analysis of effects			
Effect	df	Wald chi-square	p-value > ChiSq
Control beliefs	1	6.3117	0.0120
Self-efficacy	1	6.1388	0.0132
Rehearsal	1	5.0656	0.0244
Elaboration	1	2.7771	0.0956
Metacognition	1	4.4321	0.0353
Time study environment	1	5.2673	0.0217
Effort regulation	1	6.0998	0.0135
Help seeking	1	3.9834	0.0460
Science identity	1	13.9068	0.0002
NOS principle 1	2	5.4286	0.0663
Hypothesis revision	1	4.2310	0.0397

^aBold font indicates statistically significant values. I, information seeking actions and T, tests.

cated a participant's understanding of the tentative NOS knowledge, the more likely the participant is to perform a complex investigation (Figure 5).

Machine Learning Analysis of Investigations

Cluster analysis (Ketchen and Shook, 1996) is a machine learning method used to group sets of objects by their similarities to create a model. As clusters are added to a model, there is better coverage of the variance between objects at the cost of increased complexity. Therefore, we used the elbow method (Ketchen and Shook, 1996) to find the ideal number of clusters while minimizing error. When we examined investigations among our population, we observed three distinct clusters (Table 8). Cluster 1 can be qualitatively described as low activity, as shown by the low relative rate of engagement with all four action types, with few transitions between action types, as shown by the low mean on all bigrams that transition between action types. Cluster 2 shows increased investigative activity, particularly in regard to hypothesis generation and testing. Cluster 3 builds on high information seeking and planning. These qualitative descriptions are further visualized in Figure 2B.

We compared these cluster assignments with educational background and additional clickstream features to see whether either experience with formal biology course content or authentic science practices was related to cluster assignment. Using Fisher's exact test, we found that educational background was significantly associated with cluster assignment ($p < 0.001$, Cramer's $V = 0.27$). We noted that most non-science majors fall into either cluster 2 or 3, biology majors fall into cluster 1 or 3, and biology graduates are in clusters 2 and 3 (Table 9). This is a particularly interesting observation, because the amount of information seeking and planning present in cluster 3 would suggest that this is a more expert-like cluster.

Because performing a complex investigation is another practice associated with expert performance in SCI (Peffer and Ramezani, 2019) and was not included as a feature in our cluster analysis, a chi-square test of independence was performed to examine the relationship between complexity and cluster assignment (Table 9). The relationship between the variables was significant, $\chi^2(2, n = 123) = 27.56, p < 0.001$, and Cramer's $V = 0.45$. We noted that the majority of the investigations classified as simple were found in cluster 1. Because this cluster was also a low-activity and low-iteration cluster (meaning, the participant tended to perform multiple types of actions in a row, rather than moving between different activities), this cluster can be considered a relatively novice cluster. Cluster 2 had the highest number of complex investigations. Interestingly, cluster 3 was split fairly evenly between complex and simple investigations. This is particularly notable, because the high relative rate of information-seeking actions, extended planning at the outset of investigations, and increased testing strategy would suggest that this is a more expert-like cluster, and therefore we should see more, not equivalent, numbers of investigations that are complex in nature.

A multinomial logistic regression was fit to predict how likely individuals are to be in one of the three clusters (Table 10). Supplemental Table 6 shows more details of this logistic model. We excluded NOS principle 1 and NOS principle 2 from this model, because their correlation to the response was negatively affecting the fit of the model. This removal did not affect the fit of the model negatively, nor were those two variables important



FIGURE 5. Relationship between various predictors and associated practices by theoretical grouping. Motivation scales includes intrinsic and extrinsic goal orientation, task value, control beliefs, self-efficacy for learning, and test anxiety. Learning strategies scales include rehearsal, elaboration, organization, critical thinking, metacognition, time and study environment, effort regulation, peer learning, and help seeking. There was one scale for science identity, and the NOS assessment included two principles: principle 1 the lack of a universal scientific method; and principle 2, the tentativeness of science knowledge.

in fitting this model due to our variable selection results from random forests (Breiman, 2001). Therefore, supported by theory and statistical models, these two variables were removed from the list of the predictors, and the final model was fit.

This logistic model was significant ($\chi^2 = 89.82, p < 0.0001$) with extrinsic goal regulation, race, and gender strongly contributing to predicting the likelihood of a participant falling into any of the three categories of the cluster variable. Extrinsic goal regulation meaningfully contributed to predicting cluster 1 versus cluster 3. Although intrinsic goal regulation, task value, metacognition, time and study environment, effort regulation, and help seeking were not overall significant in predicting all three categories of clusters, all significantly contributed in predicting cluster 1 versus cluster 3. Finally, a significant difference can be observed between the category of non-science majors and biology graduates while predicting the likelihood of falling into cluster 1 versus cluster 3, meaning biology graduates were more likely to fall into cluster 1 than non-science majors were (Supplemental Table 6).

DISCUSSION

Technological and methodological advances in the field of education research such as those afforded by learning analytics have the potential to change how we conceptualize assessment. In particular, technological advances could facilitate the assessment of cognitive constructs like NOS/EBAS that cannot be

directly measured. Following best practices within the learning analytics community, namely, couching quantitative data analyses within prior qualitative work (Shaffer, 2017), this study extended our prior qualitative work and takes another important step toward the development of a rigorous, high-throughput measure of EBAS as seen through inquiry practices. Here, we expanded our model of practices as proxy for epistemological beliefs through an examination of which other cognitive factors are important for identification of the practices most likely to reflect underlying epistemological beliefs and not differences in motivation to complete the task or self-regulated learning. We also noted differences in inquiry practices among our three populations of interest, which may suggest that EBAS evolve as students complete science course work.

Development of fast quantitative assessments of NOS/EBAS will be useful in pedagogical practice to identify students with more or less sophisticated EBAS. This information could be used to both personalize learning in the classroom and evaluate new pedagogical interventions designed to improve NOS/EBAS. For example, subgroups of students who have less sophisticated EBAS about science could be targeted for additional direct instruction designed to foster the development of sophisticated epistemological beliefs. Another possible solution would be to pair students with more sophisticated EBAS with those with less sophisticated beliefs to facilitate near-peer teaching to foster development of sophisticated EBAS.

TABLE 8. Centroids for each cluster by feature average with SD in parentheses^a

Feature	Cluster 1	Cluster 2	Cluster 3
Relative rate of hypothesis generation (H)	0.15 (0.05)	0.29 (0.08)	0.13 (0.06)
Relative rate of information gathering (I)	0.02 (0.05)	0.00 (0.04)	0.52 (0.17)
Relative rate of testing (T)	0.07 (0.06)	0.45 (0.11)	0.26 (0.12)
Relative rate of concluding (C)	0.12 (0.03)	0.25 (0.06)	0.10 (0.04)
SH bigrams	0.11 (0.05)	0.25 (0.06)	0.02 (0.05)
SI bigrams	0.01 (0.03)	0.00 (0.00)	0.09 (0.05)
HH bigrams	0.00 (0.00)	0.00 (0.02)	0.00 (0.00)
HI bigrams	0.00 (0.00)	0.00 (0.00))	0.02 (0.05)
HT bigrams	0.15 (0.05)	0.28 (0.07)	0.10 (0.06)
HC bigrams	0.00 (0.02)	0.00 (0.04)	0.01 (0.02)
IH bigrams	0.01 (0.04)	0.00 (0.00)	0.09 (0.05)
II bigrams	0.00 (0.02)	0.00 (0.00)	0.38 (0.20)
IT bigrams	0.00 (0.00)	0.00 (0.00)	0.03 (0.06)
IC bigrams	0.00 (0.01)	0.00 (0.04)	0.01 (0.03)
TH bigrams	0.03 (0.05)	0.03 (0.08)	0.02 (0.04)
TI bigrams	0.00 (0.01)	0.00 (0.04)	0.02 (0.05)
TT bigrams	0.55 (0.10)	0.17 (0.17)	0.12 (0.12)
TC bigrams	0.11 (0.05)	0.24 (0.08)	0.09 (0.05)
CI bigrams	0.00 (0.00)	0.24 (0.08)	0.01 (0.03)

^aBold type indicates key dimensions for each cluster.

Because EBAS can be influenced by affective factors such as self-efficacy (Tsai *et al.*, 2011) and inquiry practices observed could be influenced by other factors such as motivation or experience with simulation content, we first set out to determine how non-epistemologically relevant factors influenced practices in SCI. We noted that, in terms of expert features identified in our prior work (Peffer and Ramezani, 2019) or in novice/expert studies in engineering (Atman *et al.*, 2007), undergraduates were similar to one another in performing fewer planning and total actions when compared with the biology graduates, who were roughly equivalent to the experts in our prior work (Figure 3). This suggests that experience with biology content does not influence the prevalence of at least these expert-like practices. We also noted that biology majors and biology graduates both did more tests than non-science majors, including when controlling for the total number of actions performed. This is somewhat different from what we observed in our previous qualitative analysis, namely, that our experts had predominantly information-seeking actions, not testing actions. This observation was reflected in our clustering analyses as well (Table 9). This could be due to differences between the ques-

TABLE 9. Clustering assignment relative to educational background, complexity, planning, and repeated actions^a

	Cluster 1	Cluster 2	Cluster 3
Non-science majors	10 (0.15)	27 (0.40)	30 (0.45)
Biology majors	20 (0.48)	8 (0.19)	14 (0.33)
Biology graduates	6 (0.43)	1 (0.07)	7 (0.50)
Simple investigation	32 (0.46)	9 (0.13)	28 (0.41)
Complex investigation	5 (0.09)	26 (0.48)	23 (0.43)
Planning actions	0.19 (0.52)	0.00 (0.00)	5.47 (5.58)
Repeated I	0.22 (0.54)	0.03 (0.17)	6.01 (5.33)
Repeated T	5.58 (2.02)	1.81 (0.82)	2.35 (1.43)

^aCounts per category are shown with relative rate in parentheses. I, information seeking actions; T, tests.

tion the students are trying to answer (e.g., Chinn and Malhotra, 2002) or updates to the interface (e.g., Quigley *et al.*, 2017). Classroom setting could also be a factor that influenced our results, as some participants used SCI during the course of their normal classroom activities, while others used the tool as part of a psychological lab study. Although this is a limitation of this work, we do note that, in general, the non-science majors' performance was equivalent to the novices described in Peffer

TABLE 10. Multinomial logistic regression analysis of effects to predict how likely individuals are to be in one of three activity-based clusters

Type 3 analysis of effects			
Effect	df	Wald chi-square	p-value > ChiSq
Intrinsic goal regulation	2	5.9406	0.0513
Extrinsic goal regulation	2	9.5016	0.0086
Task value	2	3.9817	0.1366
Control beliefs	2	1.7275	0.4216
Self-efficacy	2	1.1428	0.5647
Test anxiety	2	2.9806	0.2253
Rehearsal	2	1.1583	0.5604
Elaboration	2	0.1653	0.9207
Organization	2	0.7886	0.6742
Critical thinking	2	4.5087	0.1049
Metacognition	2	5.2514	0.0724
Time study environment	2	5.2461	0.0726
Effort regulation	2	4.7711	0.0920
Peer learning	2	3.9139	0.1413
Help seeking	2	4.8860	0.0869
Science identity	2	3.4700	0.1764
Gender	2	13.9911	0.0009
Race	4	10.8283	0.0286
Hypothesis revision	2	2.5901	0.2739
Population	4	6.0795	0.1933

and Ramezani (2019), and the biology graduates were similar to the experts in this previous work. Additional research is needed to tease out the influence of these extraneous factors on inquiry practices in SCI.

Our statistical modeling analysis revealed some surprising relationships among motivation, learning strategies, science identity, and performance on the NOS assessment (Figure 5). Figure 5 is a graphical summary of our modeling results. Directional arrows represent whether different cognitive constructs (e.g., self-efficacy, motivation) were associated with performing more novice-like (shown on the left side) or more expert-like (shown on the right side). For example, the more strongly someone identified as a science person, the more sophisticated his or her performance on NOS principle 1 (the lack of a universal scientific method), then the more planning and total actions were performed, shown as a directional arrow between science identity or NOS principle 1 and both planning and total actions (Figure 5). The increased number of actions could make sense among those who understand that there is no universal scientific method, because they may be more likely to continue the investigation until they reach a satisfactory answer, not because they followed a standard procedure of hypothesizing, testing, and concluding.

The stronger identification as a science person associated with increased activity could be reflective of confidence in being able to complete a science-related task. As a type of discourse identity (Gee, 2000), identifying with a discipline in this way suggests that one feels one belongs in a science community and feels comfortable leveraging the community's language and skills. Science identity is also influenced by an individual's interest, performance, competence, and recognition from others as a member of such a community (Hazari *et al.*, 2010). Perhaps students who identified as a science person were more likely to behave like one within the simulation.

Increased metacognitive ability and effort regulation or self-regulated learning were associated with increased number of total actions, but not necessarily planning actions (Tables 4–6 and Figure 5). This could suggest that increase in action number is more reflective of students' ability to stay on task or possibly the ability to reflect on what they know and plan what they need to do next. In future iterations of our assessment, it will be important to control for both of these factors when interpreting action number. We noted that elaboration, or the process of building internal cognitive connections between what an individual knows, was important for understanding increased planning actions. Because this phase likely includes summaries (some of which can be seen in the student's notebook), it makes sense that increased elaboration would be associated with increased planning time. Interestingly, prior work suggests planning time is a key expert practice and we also see that sophisticated beliefs on the two NOS principles assessed in this study are both associated with an increased number of planning actions. Perhaps these participants hold more sophisticated beliefs about the lack of a universal scientific method because they understand that there are different ways to test questions and sufficient preplanning is necessary to identify the best possible strategy to use. This activity could be associated with understanding of the tentativeness of science knowledge, because planning could represent getting a general idea of the current state of the field and controversies. Future work, such as

including retrospective interviews to interrogate why participants chose to plan or not, is warranted.

For complexity of investigations, we noted that biology graduates and biology majors performed predominantly more scientifically authentic, or complex, investigations, whereas the non-science majors performed predominantly simple investigations (Figure 4). Because the non-science majors are roughly equivalent to our novices in Peffer and Ramezani (2019) and the biology graduates are roughly equivalent to the experts in the same study, this result is consistent with our earlier work. We noted that increased belief in a positive outcome for a science learning experience, self-regulated learning, metacognition, and identification as a science person were associated with performing a complex investigation. Metacognition and self-regulated learning are both associated with increased performance in learning tasks (Zohar and Barzilai, 2013), and it also makes sense that individuals who believe that their performance at a task will result in a positive outcome are more likely to be engaged with the task. These non-epistemologically relevant factors will need to be controlled for during future development of our model of EBAS as seen through practices.

It is somewhat surprising to observe the converse relationship between science identity and self-efficacy. One would predict that an individual who identifies as a science person also believes in his or her ability to do well at science tasks and that both of these would predict performing a complex investigation. A sense of competence performing science tasks is a factor influencing science identity. Though identity is a complex construct, people may experience high science identity rooted in another factor (e.g., high science interest), despite not having feelings of competence when it comes to completing a science task (Carlone and Johnson, 2007). Indeed, previous work has indicated that science identity is a stronger predictor than self-efficacy of persistence into science careers (Estrada *et al.*, 2011). Our observations could indicate that students are overconfident in their ability to engage in academic science tasks, explaining the high self-efficacy score associated with performing less expert-like, simple investigations in an authentic science setting. As the self-efficacy metric used had items about students' confidence to perform academic science tasks, such as exams, it could be that academic self-efficacy is not conflated with self-efficacy to do authentic science tasks. While previous work has demonstrated the positive relationship between research self-efficacy and research skills (Adedokun *et al.*, 2013), this model does not include science identity. Future work examining context-specific self-efficacy, such as research self-efficacy, in relation to science identity and science practices may shed light on these relationships.

Higher scores on assessment of metacognitive skills in science class was associated with increased likelihood for expert-like activities, including total actions, planning actions, performing a complex investigation, and presence in cluster 3 (Table 10 and Figure 5). Metacognition and epistemic cognition (epistemological beliefs in action), alongside baseline cognitive processes like reading, are considered part of a three-part model for describing human cognition (Kitchener, 1983). In this model, the three levels of cognition build on one another, with activities such as reading at the bottom, "thinking about thinking" or metacognition in the middle, and finally "knowing about knowing" or epistemic cognition at the top (Kitchener, 1983;

Hofer, 2004). It may be possible that the increases in metacognitive ability are co-occurring with increases in epistemic cognition as well, supporting the observation that these expert-like practices are reflective of more sophisticated epistemological beliefs as seen through practices. We did not observe a large difference in metacognitive skills between any of our three populations of participants (Figure 1), which suggests that these metacognitive skills are not developing over the course of experience either with biology content or experience with authentic science practices but are instead reflective of an individual's underlying cognitive structures.

In regard to predicting cluster assignment, we looked at the likelihood ratios estimated within the multinomial logistic regression model, which was fit to predict students who are more likely to fall into one of the three clusters based on their predictor variables. The results from the Supplemental Table 6 showed students with higher intrinsic and extrinsic goal regulation, which are related to motivation, and metacognition were more likely to be in cluster 3 rather than cluster 1. Because cluster 3 was our information-seeking cluster and cluster 1 was our low-activity cluster, this suggests that students in cluster 3 generally performed more sophisticated investigations because they were both more motivated to do so and they were more reflective during the course of their investigations. In contrast, increased task value, or how interesting the student found science, was associated with assignment to cluster 1 (Table 10). This seems somewhat contradictory, because if someone is interested in science, we would predict that person would be more likely to deeply engage with the task. However, it could also mean that interests are not necessarily correlated with simulation performance, which is important to note when ascertaining how practices could reflect epistemological beliefs. Effort regulation or self-regulated learning and help seeking were also associated with assignment to cluster 1. Because increased effort regulation was also likely to be associated with increased total number of actions, it seems that the students were able to stay on task, but that their efforts did not necessarily translate into a more sophisticated investigation. This suggests that simulation practices are not necessarily due to one's ability to stay on task indicated that expert-like practice is more likely a result of differences in epistemological beliefs.

Across all of the practices that we modeled, we noted that the greatest differences were between the non-science majors and biology graduates. Although this may not be surprising, given our prior work comparing individuals with demonstrable experience with science practices (as defined by their publication in peer-reviewed literature) to individuals with no experience with authentic science practices, what is particularly notable is that the biology major practices clearly exist in between the other two populations. Sometimes the biology major practices look similar to those of their undergraduate peers, such as in total number of actions and planning actions, but at other times, their practices look more like those of the biology graduates, such as in complexity of investigations and relative rate of tests performed. This could be reflective of a progression of EBAS that occurs during the enculturation process of becoming a biologist.

Prior work comparing misconceptions about science between biology majors and non-science majors indicated that incoming non-science majors had more misconceptions about

how science works than biology majors (Cotner *et al.*, 2017). Non-science majors and biology majors also differ in terms of science identity, both at the beginning of (Cotner *et al.*, 2017) and over the course of their university careers (Figure 1). Therefore, it may not be surprising that biology major practices are more expert-like and potentially suggestive of more sophisticated underlying EBAS than the non-science majors. However, it is notable that the biology majors exist in the middle space between the biology graduates and non-science majors. This raises the question of what experiences biology majors (and later biology graduates) have, that non-science majors do not, that could potentially influence development of sophisticated EBAS. Identification of these pedagogical aspects is important for fostering development of sophisticated EBAS for all biology majors as well as for non-science majors. Because not all biology majors will pursue graduate education and non-science majors are unlikely to pursue additional science courses upon degree completion, identification of these important moments and/or pedagogies is important for fostering sophisticated EBAS and overall science literacy.

Another possible explanation for biology majors as an intermediary group could be the result of experience with biology content. One subpopulation of biology majors were enrolled in an upper-division ornithology course, the subject area of the simulation. We noted no major differences in planning actions between students in the ornithology course (mean = 1.72, SD = 2.99) and other biology majors (mean = 1.66, SD = 3.42), complexity of investigation (78% complex for ornithology students, 76% for others), and a slight increase in hypothesis revision (35% for ornithology students, 21% for other biology majors). Therefore, the differences in practices observed do not seem to be the result of familiarity with the content of the simulation. Instead, the differences we observe are likely the result of differences in underlying EBAS.

We were somewhat surprised to not observe a statistically significant difference between our populations and performance on the NOS assessment. One potential confound is that the majority of non-science majors were students in the lead author's nonmajors biology course, and she strongly emphasized NOS principles in class. This is particularly relevant, because direct instruction is a known pedagogical best practice for improving student NOS understanding (Khishfe and Abd-El-Khalick, 2002). We also noted when scoring the open-ended response items that one student commented "As stated in Dr. Peffer's lecture, scientific investigations can follow more than one method as science isn't just followed by a linear pathway." Therefore, it is possible that the non-science majors were scoring higher than would be expected.

Although this is a potential confound and limitation of this study, it is interesting to note that the distinction between a sophisticated (1) and naïve (3) score on this assessment was important for predicting practices such as complexity of investigation and planning actions, as well as the number of non-science majors present in the more expert-like cluster, cluster 3 (Table 9). It may be that the students who received the direct instruction did better on both the NOS assessment and had correspondingly more expert-like investigations. When looking more generally across all practices, non-science major practices were roughly equivalent to those of the novices in Peffer and Ramezani (2019). This novice population was also exclusively

non-science majors enrolled at a different institution, none of whom were taking a biology course taught by the lead author. This suggests that, although presence in the lead author's class could have influenced non-science majors' performance on the NOS metric (and potentially investigative approach), overall practices among this group appear to be generalizable across two institutions. Furthermore, the novices in Peffer and Ramezani (2019) were almost exclusively in their fourth and final year of their university studies, whereas the population studied here was more diverse in terms of degree progress. This suggests that years of schooling does not influence practices or EBAS/NOS understanding among non-science majors.

To extend our previous qualitative analysis of differences in inquiry practices, we used the log files generated by individual participants as they engaged in SCI to perform *k*-means clustering. We observed three emergent clusters (Figure 2B). When comparing the clusters with educational background and complexity of investigation (another marker of expert-like practices and potential hallmark of a more sophisticated epistemology), we noted that cluster 3 appeared to be more expert-like. Cluster 3 was characterized as having high information-seeking and planning activity. Because we see the highest number of non-science majors and biology graduates in this cluster, and the second highest number of biology majors, it suggests that experience with biology course work or progress to degree completion is unrelated to a more expert-like investigative style in SCI. Regarding complexity of investigation, we noted that the majority of simple investigations were found in the low-activity cluster, cluster 1, which appeared reminiscent of the novice practices in Peffer and Ramezani (2019). Interestingly, we noted that cluster 2 and not cluster 3 had the highest number of complex investigations (Table 9). We also noted that cluster 3 contained approximately half simple and complex investigations. Given the other expert-like practices in this cluster, this is a surprising observation. Future work will examine how participants in cluster 2 The high amount of testing by participants in cluster 2 used the information they collected, either from testing or looking for information they collected, either from testing or looking outside the simulation, are warranted to better understanding how practices relate to sophistication of investigation and overall NOS/EBAS. We also noted that both populations of undergraduates were more likely to fall into either cluster 1 or cluster 2, rather than cluster 3, which may also suggest that students with investigations similar to those in cluster 3 may also have more sophisticated EBAS.

Although we observed no statistically significant differences between cluster assignments among our biology graduate population, we did note qualitatively that of the seven biology graduates assigned to cluster 3, four were postdoctoral associates, two were doctoral candidates, and one was a master's student. Within cluster 1, half of the graduates' group were master's students, two were doctoral students, and one was a postdoctoral associate. We also noted that postdoctoral associates performed relatively less testing when accounting for the length of their investigations than the other two populations and more information-seeking actions, particularly in the planning phase before beginning their investigations. Because the postdoctoral associates were most similar to the expert population used in Peffer and Ramezani (2019), this lends support to cluster 3 as our expert-like cluster.

CONCLUSIONS

Developing an assessment of EBAS/NOS is a challenging yet important task for improving student outcomes in all science classes. A better assessment of EBAS/NOS will enhance understanding of how EBAS develop in the classroom and will be useful for developing evidence-based pedagogical strategies that can be leveraged to ultimately lead to improved pedagogy and science literacy for both non-science and science majors. This work contributes to a growing literature that supports the use of technology and learning analytics to assess latent constructs such as EBAS. Examination of practices in an authentic science activity like inquiry provides new insights into how students conceptualize NOS knowledge, which is valuable information for both instructors and researchers. Using technology allows for high-throughput and fast access to this information and ease in gathering data in real time during a course. This facilitates just-in-time teaching and improved not only course outcomes, but overall science literacy.

We note that, in the context of biology learning and teaching, this study suggests that biology content knowledge may need to be considered separately from epistemological beliefs as seen through inquiry. We also see biology major practices and potentially EBAS existing in an intermediary zone between biology graduates and non-science majors. This is an important consideration when determining benchmarks for what knowledge and skills are necessary for students enrolled in biology programs, and how beliefs about the NOS knowledge develop over the course of completing a university degree.

REFERENCES

- Abd-El-Khalick, F. (2012). Examining the sources for our understandings about science: Enduring confluences and critical issues in research on nature of science in science education. *International Journal of Science Education, 34*(3), 353–374.
- Adedokun, O. A., Bessenbacher, A. B., Parker, L. C., Kirkham, L. L., & Burgess, W. D. (2013). Research skills and STEM undergraduate research students' aspirations for research careers: Mediating effects of research self-efficacy. *Journal of Research in Science Teaching, 50*(8), 940–951. doi: 10.1002/tea.21102
- Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: "Views on Science—Technology—Society" (VOSTS). *Science Education, 76*(5), 477–491.
- 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
- Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
- Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187–1218. doi: 10.1002/tea.20237
- 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
- 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.
- Cotner, S., Thompson, S., & Wright, R. (2017). Do biology majors really differ from non-STEM majors? *CBE—Life Sciences Education, 16*(3), ar48.
- Cribbs, J. D., Hazari, Z., Sonnert, G., & Sadler, P. M. (2015). Establishing an explanatory model for mathematics identity. *Child Development, 86*(4), 1048–1062. doi: 10.1111/cdev.12363
- Crowell, A. J., & Schunn, C. D. (2016). Unpacking the relationship between science education and applied scientific literacy. *Research in Science Education, 46*(1), 129–140. https://doi.org/10.1007/s11165-015-9462-1

- 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>
- Dogan, N., & Abd-El-Khalick, F. (2008). Turkish grade 10 students' and science teachers' conceptions of nature of science: A national study. *Journal of Research in Science Teaching, 45*(10), 1083–1112.
- Elby, A., Macrander, C., & Hammer, D. (2016). Epistemic cognition in science. In Bråten, I., Greene, J., & Sandoval, W. (Eds.), *Handbook of epistemic cognition* (pp. 113–127). New York: Routledge.
- Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology, 103*(1), 206–222. doi: 10.1037/a0020743
- Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2016). Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice. *Journal of Engineering Education, 105*(2), 312–340. doi: 10.1002/jee.20118
- Gee, J. P. (2000). Chapter 3: Identity as an analytic lens for research in education. *Review of Research in Education, 25*(1), 99–125.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations, 11*(1), 10–18.
- Hazari, Z., Sonnett, G., Sadler, P. M., & Shanahan, M. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching, 47*(8), 978–1003. doi: 10.1002/tea.20363
- Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist, 39*(1), 43–55. doi: 10.1207/s15326985ep3901_5
- 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>
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology and Society, 15*(1), 38–52.
- Joint Information Systems Committee. (2015). *Code of practice for learning analytics*. Published under the CC BY 4.0 license.
- Ketchen, D., & Shook, C. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*(6), 441–458.
- Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching, 39*, 551–578.
- Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition. *Human Development, 26*(4), 222–232.
- 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.
- Kukreja, S. L., Löfberg, J., & Brenner, M. J. (2006). A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. *IFAC Proceedings, 39*(1), 814–819.
- 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>
- Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology, 71*(5), 353.
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439–1459.
- Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16*(5), 492–509.
- McComas, W. F. (2015). The nature of science & the next generation of biology education. *American Biology Teacher, 77*(7), 485–491. <https://doi.org/10.1525/abt.2015.77.7.2>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.
- Noble, W. S. (2009). How does multiple testing correction work? *Nature Biotechnology, 27*(12), 1135–1137.
- Organisation for Economic Co-operation and Development. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris.
- Peffer, M. E., Beckler, M., Schunn, C., Renken, M., & Revak, A. (2015). Science classroom inquiry (SCI): A novel simulation to scaffold science learning. *PLoS ONE, 10*(3). <https://doi.org/10.1371/journal.pone.0120638>
- Peffer, M. E., & Kyle, K. (2017, March). 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).
- Peffer, M. E., Quigley, D., & Mostowfi, M. (2019, March). Clustering analysis reveals authentic science inquiry trajectories among undergraduates. In *Proceedings of the ninth international conference on learning analytics & knowledge* (pp. 96–100). <https://doi.org/10.1145/3303772.3303831>
- Peffer, M. E., & Ramezani, N. (2019). Assessing epistemological beliefs of experts and novices via practices in authentic science inquiry. *International Journal of STEM Education, 6*(1). <https://doi.org/10.1186/s40594-018-0157-9>
- Peffer, M. E., & Renken, M. (2015, June). Science classroom inquiry (SCI) simulations for generating group-level learner profiles. In *Exploring the material conditions of learning: The computer supported collaborative learning (CSCL) conference 2015* (Vol. 2, pp. 707–708).
- Peffer, M. E., Roysse, E., & Abelein, H. (2018, June). Influence of affective factors on practices in simulated authentic science inquiry. In *Rethinking learning in the digital age: Making the learning sciences count, 13th international conference of the learning sciences (ICLS)* (Vol. 2, pp. 997–1000).
- Pintrich, P. R. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.
- Quigley, D., Ostwald, J., & Sumner, T. (2017, March). Scientific modeling: Using learning analytics to examine student practices and classroom variation. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 329–338). <https://doi.org/10.1145/3027385.3027420>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- 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, NC: 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>
- 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>
- Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press, Lulu.com.
- Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In *Handbook of research on educational communications and technology* (pp. 311–321). New York: Springer.
- Tsai, C. C., Ho, H. N. J., Liang, J. C., & Lin, H. M. (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.

- Wang, J., & Hazari, Z. (2018). Promoting high school students' physics identity through explicit and implicit recognition. *Physical Review Physics Education Research*, 14(2), 020111. <https://doi.org/10.1103/PhysRevPhysEducRes.14.020111>
- Wong, S. L., & Hodson, D. (2009). From the horse's mouth: What scientists say about scientific investigation and scientific knowledge. *Science Education*, 93(1), 109–130. <https://doi.org/10.1002/sce.20290>
- Wong, S. L., & Hodson, D. (2010). More from the horse's mouth: What scientists say about science as a social practice. *International Journal of Science Education*, 32(11), 1431–1463. <https://doi.org/10.1080/09500690903104465>
- Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: Current and future directions. *Studies in Science Education*, 49(2), 121–169.