Assessment of Language in Authentic Science Inquiry Reveals Putative Differences in Epistemology

Melanie E. Pepper, PhD
University of Northern Colorado
Ross Hall 1556 Campus Box 92
Greeley, CO 80639
+1 970-351-2923
melanie.pepper@unco.edu

Kristopher Kyle, PhD
University of Hawai'i at Manoa
587 Moore Hall, 1890 East-West Rd
Honolulu, HI 96822
+1 808-956-8610
kkyle@hawaii.edu

ABSTRACT

Science epistemology, or beliefs about what it means to do science and how science knowledge is generated, is an integral part of authentic science inquiry. Although the development of a sophisticated science epistemology is critical for attaining science literacy, epistemology remains an elusive construct to precisely and quantitatively evaluate. Previous work has suggested that analysis of student practices in science inquiry, such as their use of language, may be reflective of their underlying epistemologies. Here we describe the usage of a learning analytics tool, TAALES, and keyness analysis to analyze the concluding statements made by students at the end of a computer-based authentic science inquiry experience. Preliminary results indicate that linguistic analysis reveals differences in domain-general lexical sophistication and in domain-specific verb usage that are consistent with the expertise level of the participant. For example, experts tend to use more hedging language such as “may” and “support” during conclusions whereas novices use stronger language such as “cause.” Using these differences, a simple, rule-based prediction algorithm with LOOCV achieved prediction accuracies of greater than 80%. These data underscore the potential for the use of learning analytics in simulated authentic inquiry to provide a novel and valuable method of assessing inquiry practices and related epistemologies.

CCS Concepts

• Applied Computing □ Education □ Interactive Learning Environments.

Keywords

Science epistemology; Authentic Science Inquiry; Science Classroom Inquiry Simulations; TAALES; Science Practices; Lexical sophistication; Assessment

1. INTRODUCTION

1.1 SCI and Authentic Inquiry

Practicing scientists perform authentic science inquiry [1]. Unlike

simple inquiry, authentic inquiry is typically non-linear and involves a variety of complex features (e.g., proposing mechanisms, performing multiple studies, revising hypotheses) [1-2]. Teachers most often employ simple inquiry in classrooms due to concerns about safety, time, expense and need for individualized scaffolding [1-3]. Although simple inquiry may be easier to perform in the classroom, a lack of exposure to authentic science inquiry results in students developing an understanding of how science works that is inconsistent with real world science practices [2].

Science Classroom Inquiry (SCI) simulations were developed to address pedagogical and research constraints associated with authentic science inquiry experiences for students [3]. SCI simulations position students as researchers during a computer-based simulated authentic inquiry activity. We demonstrated in previous work that after students engage in SCI, their understanding of what it means to do science is altered to a more sophisticated stance [3]. On post-intervention metrics, students commented that completing a simulation helped them realize science is much more complicated than they originally thought. They also reported a new understanding that problems do not always have a single correct answer. We noted that when students are given the option to engage in non-linear, authentic science inquiry, their practices are diverse [2]. Diverse inquiry practices, such as running multiple investigations, seeking additional information, and coordinating evidence with theory, may provide insight into students' epistemological beliefs about science.

1.2 Epistemology in Authentic Science Inquiry

An individual's science epistemology encompasses a set of beliefs about the nature of science and science knowledge that an individual possesses. Although attaining more sophisticated or expert-like epistemological beliefs is widely recognized as important for science literacy [4], epistemology is difficult to precisely define and consequently measure. For example, current pen and paper assessments operate under the assumption that the user is interpreting the questions the same as the researcher [5]. Furthermore, epistemological beliefs about science can vary widely both within and outside of disciplines. Consequently, there is no definite consensus about what a “correct” answer would be on an assessment. Studying student practices in authentic inquiry, particularly via the artifacts generated, discourse surrounding these artifacts, and interrogation into decisions made, is likely to provide markers of an individual's science epistemology [5]. Furthermore, student performance at particular points during authentic inquiry and examination of trends over certain tasks or periods of time may indicate which specific epistemological

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
LAK'17, March 13-17, 2017, Vancouver, BC, Canada.
Copyright is held by the owners/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-4870-6/17/03...\$15.00

DOI: <http://dx.doi.org/10.1145/3027385.3027425>

beliefs or constructs (e.g., the tenuous nature of science, lack of a universal scientific method) are engaged during inquiry [5].

1.3 Learning Analytics for Assessment of Epistemology

Learning analytic methods embedded in simulated authentic science inquiry are a possible solution for assessing epistemological beliefs about science [6]. The value of using learning analytics to capture data generated in real time by participants while engaged in authentic inquiry provides researchers and teachers with greater insight into the participant's underlying epistemological beliefs. Furthermore, learning analytics allows for capture of data in an authentic environment, such as during a SCI simulation. We use learning analytics here in an apprenticeship manner [6] to assess word choice between experts and novices during the conclusion generating segment of a SCI simulation. Previous work has suggested that language used during engagement in scientific practices may be reflective of epistemology, particularly as discourse around a topic is associated with the process of making sense about a task [6-7]. Therefore, we used both domain-general and domain-specific indices of linguistic sophistication. Domain-general indices included content word frequency calculated by the Tool for the Automatic Analysis of Lexical Sophistication (TAALES) [8]. TAALES has been used in a number of studies to measure lexical sophistication in a number of domains (e.g., writing quality [9], lexical development [10], and sentiment analysis [11].) Domain-specific indices were calculated to determine specific linguistic items associated with expert and novice writing using natural language processing (NLP) technology and methods based in corpus linguistics [12].

1.4 Current Study

Given the limitations with current pencil and paper assessment of epistemological beliefs about science, we sought to determine a method for using learning analytic techniques to more quickly and accurately assess epistemology in an authentic environment, namely during the authentic science inquiry experience provided by SCI simulations. Here we describe our preliminary results in which we examined the language used by novices and experts when making conclusions at the completion of authentic science inquiry. We discuss implications of using learning analytics in simulated authentic science inquiry environments, like SCI, to produce more fine-grained and robust assessments of science epistemology, as well as their pedagogical potential.

2. METHODOLOGY

2.1 Participants

20 novices and 8 experts from a large southeastern United States city participated in this study. Experts and novices were distinguished by their experience in authentic science practices, not in their experience with the simulation subject matter. For example, experts had engaged in independent biology research for at least 2 years and all were listed as authors on journal publications submitted or published at the time of the study. Novices had little if any experience in authentic science inquiry, and none were authors on primary research manuscripts. Ethnic backgrounds of participants were diverse. The novice population was comprised of 15% White/European-American, 45% Black/African-American, 15% Asian-American, 15% Indian-American, 5% multi-racial, and 5% other. The expert population was comprised of 62% White/European-American and 38% Asian-American. Among both novices and experts, participants were predominately female (novices: 69% female, 31% male;

experts: 88% female, 12% male). All novice participants had completed at least one year of college education, with the majority being in their third or fourth year of college (70% seniors, 20% juniors and 10% sophomores) and experts had passed qualifying exams to advance to doctoral candidacy or had completed doctoral training.

2.2 Data Collection

Data was collected over a single meeting in several forms. Participants were logged into the *Unusual Mortality Events* SCI simulation [6]. As the students completed the simulation, the SCI engine recorded their activity. These logs included both student generated notes and responses to simulation prompts. For the current study, we focused on the conclusion sections of the notebooks, which comprised of responses to the following two questions: 1. What is/are your final conclusion(s)? 2. What evidence supports these conclusion(s)? The average length of response for participants was 85.89 words (SD = 49.22). Experts used an average of 107.88 words (SD = 77.11, total words analyzed 863) in their responses and novices an average of 85.89 (SD = 49.22, total words analyzed 1542).

2.3 Data Analysis

The domain-general analysis was conducted using TAALES 2.0. We selected content word (i.e., nouns, verbs, adjectives, and adverbs) frequency based on the academic section of the Corpus of Contemporary American English (COCA). COCA is a large (450-million word) corpus of language samples collected in the United States since 1990. The academic section of COCA represents "general" academic language. Content word frequency has been used successfully to model lexical sophistication in a number of studies [8].

The domain-specific analysis was conducted using a method related to keyness analysis. Keyness analysis involves identifying linguistic items (usually words) that occur statistically significantly more often in one set of texts than another. We conducted an analysis that is conceptually similar to keyness analysis, though our dataset was not large enough to warrant statistical comparisons. The analysis proceeded as follows. First, all texts were tagged for part of speech by Stanford CoreNLP [13-14]. Lemmatized verb type lists were then extracted from the novice and expert texts, resulting in verb use profiles for each text type that indicated the percentage of texts each verb occurred in. The expert and verb texts were then compared. Any verb that occurred in at least two expert texts and occurred at least 20% more often in expert texts than in novice texts were considered "expert" verbs. Any verbs that met the same criteria with regard to novice texts were considered "novice" verbs. Individual texts were then reanalyzed for their use of domain-specific "expert" and "novice" verbs.

Simple, rule-based algorithms were then used with leave one out cross-validation (LOOCV) to predict whether a text should be classified as "expert" or "novice" using one of three predictors: average content word frequency, "expert" verb use, and "novice" verb use.

3. RESULTS

Since understanding science knowledge as tenuous is generally considered a sophisticated epistemological stance [7], we decided to examine language used during the conclusion phase of authentic science inquiry. We would predict that someone with less sophisticated epistemological beliefs about the tenuous nature of science would be more likely to use concrete language such as

“prove” “right” or “correct” when making their conclusions and less hedging language such as “may” “could” or “support.” To this end our approach included two levels of analysis (1) domain-general lexical sophistication as analyzed using TAALES and (2) domain-specific language between experts and novices.

3.1 Domain-general lexical sophistication

The results with regard to TAALES content word frequency suggested that experts tend to use content words that are less frequent ($M = 608.17$, $SD = 251.76$) in normal academic English use than novices ($M = 953.25$, $SD = 311.64$) (Figure 1).

Figure 1. Experts use less frequent content words.

A simple, rule-based prediction algorithm with LOOCV achieved a prediction accuracy of 82.1%. Table 1 comprises the model confusion matrix, which compares actual and predicted group membership.

Table 1. Confusion matrix for domain-general model.

	predicted expert	predicted novice	accuracy
actual expert	6	2	75.0%
actual novice	2	20	90.0%
overall accuracy			82.1%

3.2 Domain-specific verb sophistication

Table 2 comprises the results from the verb-use comparison analysis. Expert verb-use tended to be more uniformly distinct from novice verb use as indicated by the number of expert verbs versus the number of novice verbs.

Table 2. Verb lists for expert and novice participants.

Expert Verbs	Novice Verbs
result	cause
may	die
do	
support	
lead	
change	
increase	
decrease	
look	
make	

3.2.1 Expert verb use

As expected, conclusions written by experts tended to include more expert verbs ($M = 3.38$, $SD = 2.56$) than conclusions written by novices ($M = 0.85$, $SD = 0.75$) (Figure 2).

Figure 2. Use of expert verbs.

A simple, rule-based prediction algorithm with LOOCV achieved a prediction accuracy of 89.3%. Table 3 comprises the model confusion matrix, which compares actual and predicted group membership.

Table 3. Confusion matrix for the expert verb model.

	predicted expert	predicted novice	accuracy
actual expert	5	3	62.5%
actual novice	0	20	100%
overall accuracy			89.3%

3.2.2 Novice verb use

As expected, conclusions written by novices tended to include more novice verbs ($M = 0.95$, $SD = 0.52$) than conclusions written by experts ($M = 0.38$, $SD = 0.52$) (Figure 3).

Figure 3. Use of Novice verbs.

A simple, rule-based prediction algorithm with LOOCV achieved a prediction accuracy of 71.4%. Table 4 comprises the model confusion matrix, which compares actual and predicted group membership.

Table 4. Confusion matrix for the novice verb model.

	predicted expert	predicted novice	accuracy
actual expert	5	3	62.5%

actual novice	5	15	75.0%
overall accuracy			71.4%

4. DISCUSSION

4.1 Experts Differ from Novices in Lexical Sophistication

In this paper, we performed a preliminary analysis of the language used by novices and experts during authentic science inquiry using a learning analytics tool, TAALES, and keyness analysis. Our goal was to assess if use of learning analytics tools such as TAALES could be useful in assessing scientific epistemological beliefs in a more practical, high-throughput, and sensitive manner than existing pen and paper assessments. The results indicate that both domain-general and domain-specific linguistic features can be used to accurately classify expert and novice investigation conclusions, although domain-specific indices may be more construct relevant. These results are discussed below.

4.1.1 Expert v. Novice Content Word Usage

Analysis of content words used during the concluding phase indicated that experts used words that are less frequent in normal academic English than novices. Although this was predictive of expertise, it may be more closely related to overall language ability. Since experts had spent considerably more time than novices in an academic environment, it is probable that this is a reflection of their experience in academic settings. Thus, while content word frequency demonstrates predictive validity, it may be construct irrelevant. Content word frequency may be an artifact of the expert's overall language abilities, not an indicator of underlying epistemologies.

4.1.1.1 Expert v. Novice Verb Usage

Analysis of verbs used by experts and novices resulted in two independent verb lists (Table 2). Ten verbs met the criteria of occurring 20% more often in expert texts than novice texts, while only two verbs met the criteria with regard to novice texts. This suggests that experts used a more cohesive set of verbs than novices. This provides preliminary evidence that a) there is an accepted repertoire of verbs used by expert scientists to discuss research findings and b) novices have not learned to use these verbs in this setting. Semantically, many of the expert verbs represented hedging and tentative language. For example, we observed usage of the words “may” and “support” which is consistent with our initial hypothesis. See Table 5 for examples of tentative and hedging verbs in expert texts. This word choice may be reflective of an understanding of the tentative nature of science.

Table 5. Examples of tentative and hedging verbs in expert texts.

<i>This may explain the loss of this particular species from the estuary.</i>
<i>I observed a temperature change which supports my idea.</i>
<i>Specifically, the growth and ingestion of <i>Gracilaria</i> by Manatees and the loss of food items for Dolphins and Pelicans which may have made them more susceptible to infection.</i>

Importantly, using the presence of “expert verbs” as a variable in a prediction model was successful and achieved 89.3% classification accuracy. No novice texts were misclassified as expert, but 37.5% of expert texts were misclassified. This suggests that novice scientists rarely use expert verbs when explaining their

findings, but experts also avoid using these verbs occasionally. Qualitative analysis suggests that experts may also use linguistic features other than verbs to hedge and express tentativeness. For example, one of the expert conclusions included no expert verbs. Instead, it included adverbial phrases that express tentativeness such as in the sentence *Pelican **most likely** died from the parasite*. This is a potentially fruitful avenue for future research.

4.2 Word Choice May Reflect Epistemology

Since science knowledge is generally accepted as being tentative and subject to change pending new results and discoveries, the majority of scientists tend to use tentative language to speak about their results. Therefore, the use of tentative language may be reflective of the epistemological belief that scientific knowledge is not concrete. An alternative interpretation of this observation could be that experts have simply learned the language of science, in which their epistemologies may or may not be reflected. It may also be the case that learning scientific language is an important stepping stone on the way to attaining more sophisticated epistemological beliefs. Future work that ties an analysis of lexical sophistication and verb usage to existing metrics of epistemology and/or nature of science understanding (nature of science meaning the set of generally accepted facts that influence one's epistemological beliefs), may be useful for better understanding how language choices relate to epistemology. Furthermore, it may be the case that word choice is only one part of a larger investigation and the greater context in which words are utilized is more indicative of epistemology. For example, if we observe that experts are using more tentative language during their concluding phases, is this also apparent throughout their scientific investigations? Perhaps an understanding of the tentativeness of science is also reflected in the number or types of tests a user may perform when engaged in authentic science inquiry. A user who uses more definitive language may also perform simple inquiry, only performing one or two tests and concluding with a final “correct” answer. Conversely, an expert may spend more time doing varied tests before coming up with several possible conclusions that they discuss with tentative language. Or, perhaps a user that receives conflicting or inconclusive test results is more likely to use tentative language. There may also be domain-specific variations in word choice. Future work is necessary to determine how different practices during authentic science inquiry relate to language, and how these different aspects of inquiry combine to reveal insights into the user's epistemological beliefs.

4.3 Limitations

As stated above, the use of tentative language may be reflective of underlying epistemological beliefs about the tentative nature of science. However, epistemological beliefs about science include more than an understanding of the tentative nature of science knowledge. For example, other important beliefs include how science knowledge is justified and produced, the lack of a universal scientific method, and the sources of scientific information. Epistemologies are also said to be situated, meaning that an individual's epistemology will vary depending on if the student is performing inquiry in a classroom, an informal setting, or in a psychological lab study, such as in this study. Furthermore, epistemological beliefs vary widely both across and within disciplines. Therefore, although language may be a useful proxy of one of these factors, it may only be providing a very small sliver of a larger picture. Future studies examining practices in a variety of settings will be useful for determining the key epistemological episodes by setting.

Our data set for this study is small, particularly among the expert population. For example, normal sample sizes for TAALES analysis would be 100-250 participants, with keyness analysis using even larger data sets. We also note the number of words collected per participant is also small. However, given that we are seeing trends that reflect what we expect to see regarding novice/expert language usage and that the LOOCV model is predictive of expertise, we feel that a scale-up of this study is justified and may be informative.

In addition to a small data set, research participants were predominantly female in both the expert and novice populations. Since previous work on gender differences in science epistemology are conflicting, with some reporting an effect of gender and others not observing gender differences [15], this may indicate that the observed trends may change with additional male participants. However, since we are comparing two predominantly female data sets together, it is likely that the expertise differences we observe will be maintained.

4.4 Implications and Future Directions

Current pen and paper assessments of epistemology are limited. Learning analytic techniques, such as TAALES, or TAALES in combination with other methods of automatically capturing critical data as students engage in inquiry may be a new way of assessing amorphous constructs such as epistemology. Tracking students as they engage in a completely autonomous inquiry experience provides a level of authenticity that is not provided through pen and paper assessment. In addition, there is more freedom for participants to perform many different kinds of investigation, rather than limiting their responses to pre-defined survey items. This removes a common limitation cited by those who critique epistemology or nature of science assessments, namely that these metrics are based on the assumption that the user is interpreting the questions the same as the researcher [5].

In addition to utility for assessing of epistemology by researchers, use of learning analytics techniques in a simulation environment may also have important roles in instruction and pedagogy. For example, learning analytics methods could be used to provide critical feedback to teachers and/or users for instruction. [6] points out that learning analytics for assessing epistemology can be embedded in instruction, thereby giving the user real-time feedback and prompting metacognitive reflection. Alternatively, identifying important features of epistemology via learning analytic methods embedded into simulations may be useful for instructors to identify where student needs are and tailor instruction appropriately.

Although there is significant potential with learning analytics techniques for assessment of epistemology, it is necessary to consider the information garnered via learning analytics within the greater context of the learning environment. As cautioned by [7], what is chosen to be assessed and how it is interpreted is a factor of the views and biases of the researcher/teacher. This raises particularly important questions in regards to epistemology since beliefs about the nature of science and science knowledge ranges widely between scientists, even within the same discipline. In spite of these caveats, the research potential of learning analytics in the context of authentic inquiry is rich and exciting.

5. ACKNOWLEDGMENTS

This project was supported through a data consortium fellowship (NSF 112-1549112). We thank Mike Tissenbaum and Matthew Berland for choosing our project to receive a fellowship. We

thank Don Davis and the Sona team for support and use of the Counseling and Psychological Services Research Participant System. We thank Merrin Oliver for her help with data collection.

6. REFERENCES

- [1] 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.
- [2] Peffer, M. E., & Renken, M. 2015. Science Classroom Inquiry (SCI) Simulations for Generating Group-Level Learner Profiles. In *Exploring the Material Conditions of Learning: the CSCL Conference 2015*, 707-708.
- [3] Peffer, M. E., Beckler, M. L., Schunn, C., Renken, M., & Revak, A. 2015. Science Classroom Inquiry (SCI) Simulations: A Novel Method to Scaffold Science Learning. *PloS one*, 10, 3. e0120638.
- [4] Sandoval, W. 2014. Science education's need for a theory of epistemological development. *Science Education*, 98, 3. 383-387.
- [5] Sandoval, W. A. 2005. Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, 89, 4. 634-656.
- [6] 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.
- [7] Deng, F., Chen, D.T., Tsai, C.C., & Chai, C.S. 2011. Students' Views of the Nature of Science: A critical review of research. *Science Education*, 95, 6. 961-999
- [8] Kyle, K. & Crossley, S.A. 2015. Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49, 4. 757-786
- [9] Kyle, K. & Crossley, S.A. 2016. The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 12. 12-24
- [10] Crossley, S., Kyle, K., & Salsbury, T. 2016. A usage-based investigation of L2 lexical acquisition: The role of input and output. *The Modern Language Journal*, 100, 3. 702-715.
- [11] Skalicky, S., & Crossley, S. 2015. A statistical analysis of satirical Amazon.com product reviews. *The European Journal of Humour Research*, 2, 3. 66-85.
- [12] Scott, M. & Tribble, C., 2006. *Textual Patterns: Keyword and corpus analysis in language education*. John Benjamins Publishing Company, Amsterdam, The Netherlands.
- [13] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., & McClosky, D. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 55-60
- [14] Toutanova, K., Klein, D., Manning, C., & Singer, Y. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003*. 252-259.
- [15] Conley, A., Pintrich, P., Vekiri, I., & Harrison, D. 2004. Changes in epistemological beliefs in elementary science students. *Contemporary educational psychology*, 29, 2. 186-204.