Investigating Curiosity in Student Text Data

Paul Meisner, Mitchell Hanson, Naeem Seliya,
Benjamin Fine, Rushit Dave*, and Mounika Vanamala

Computer Science Department
University of Wisconsin – Eau Claire,
Eau Claire, WI 54701
seliyana@uwec.edu

Abstract

We present a unique question-based student text responses analysis that can help instructors better identify what drives students to be more engaged in their learning. To determine the level of inquisitiveness among students, data is collected utilizing the Question Formation Technique. Data collection involves presenting students with thought-provoking QFocus statements, prompting them to formulate their responses in form of questions. The data is analyzed through Natural Language Processing, which is then analyzed using the WEKA machine learning tool. Feature selection is performed using filter-based feature rankers and wrapper-based feature subset algorithms. The course subject instructors determined that the extracted features provide meaningful insight into the "Propensity for Exploration" within the student text responses as a measure of their curiosity. Through an empirical mining of words/sentences that prove a curious disposition in text data produced by students in response to thought-provoking and critical thinking analysis, we obtained promising results, including an interesting distribution of results among the different applied feature ranker and subset algorithms.

* Department of Computer Information Science, Minnesota State University, Mankato.

1. Introduction

The skill of how to learn and apply new knowledge is a vital skill students need to develop. A student's curiosity in exploring a topic supports learning that knowledge [1], building upon what is taught in the given course. Curiosity has been associated with workplace learning and job performance [3]. Curiosity supports lifelong learning, a desirable outcome of students' education [2]. Given the benefits curiosity can have on self-directed learning and job performance, it is important to be able to identify whether students are exhibiting curiosity in the assignments and lab work.

Text mining has seen increasing focus on the investigation of sentiment [4], behavior analytics [5], linguistic understanding [6] improving product marketing [7], and pedagogical improvements [8]. Our project focuses on a relatively novel area, i.e., curiosity detection in text. This paper presents preliminary, yet promising, results of empirically mining words that demonstrate a curious disposition (of the students) in text data produced by students in response to thought-provoking and critical-thinking exercises. The success of our project could positively impact efforts to assess both curiosity and its impact on educational outcomes.

Grossnickle [9] has provided a framework for understanding facets, factors, and dimensions of the construct of curiosity that are relevant to the education audience. The key dimensions identified in the framework for curiosity include focus of curiosity (physical, perceptual, social, and epistemic), scope of curiosity (breadth vs. depth), cause of curiosity (diversive vs. specific or interest vs. deprivation), and consistency of curiosity across situational contexts (state vs. trait) [9]. Curiosity is positively linked to inquiry-based learning [10]. Questions are artifacts of curiosity. People consider children as curious when they ask many questions about a variety of topics, and particularly when they creatively combine ideas. Questions are posed to bring to light that which is unknown or not fully formed within the mind of the individual posing the question.

It seems reasonable to consider question data sets (especially from students) as a starting point for detecting a curious disposition. The datasets for our data mining approach to curiosity detection in students' text come from applying the Question Formulation Technique (QFT) [11], originally developed by the Right Question Institute [12], in an upper-level Artificial Intelligence (AI) course in an undergraduate computer science program. Previously, we performed relatively similar work for a lower-level Electric Circuits course in an undergraduate Electrical Engineering program. A portion of five class sessions in the AI course was utilized to obtain the QFT data, to improve students' ability to formulate questions and to support their curiosity on course topics. Compared to an expert examining student text data to determine curiosity levels, we envision our data mining solutions could provide substantial aid to experts. The solutions developed will be useful to detect whether curiosity is demonstrated in the results of the QFT exercises, provide analysis on key dimensions of curiosity, and potentially predict associated behaviors of students'.

The metric used in our study to assess the curiosity level of each student's question in the QFT data is "Propensity for Exploration". This metric is chosen because the dimension of curiosity that is most relevant to self-directed learning is the desire to identify knowledge gaps and seek out knowledge to close those gaps [14]. Propensity for exploration (PE) attempts to capture the identification of knowledge gaps and demonstration of some understanding of the landscape of the topic, which supports curiosity and the desire to seek out the knowledge [14]. Specifically, PE considers the degree to which the question identifies characteristics of, or layers within, the subject of the question, the degree to which relationships between the primary subject and other topics is identified, how relevant those characteristics and relationships are, and how well the question directs the attention of the audience within the landscape of the topic. Each question in the dataset is labeled as belonging to one of two categories for PE: 1 (Low) and 2 (High).

In a given dataset (student text responses set as per the QFT process), all unique words (tokens) are considered as features or attributes, after removing general stop-words typically observed in text data. Feature Selection (FS) methods have been applied to reduce the high dimensionality of the obtained datasets. We investigate five different FS and they include: two wrapper-based feature subset selection methods and three filter-based feature ranker techniques. The wrappers involve the C4.5 decision tree classifier with the BestFirst and GreedyStepwise search algorithms, while the filters consist of the ChiSquared, ReliefF, and GainRatio algorithms. The algorithm and associated parameter details for the C4.5 classifier and the five feature selection methods considered in our study is provided in [15]. Each FS method provides a reduced set of features for domain experts to examine to determine whether the selected features are indeed correlated with the different PE levels.

The important conclusions determined from our case study are that the two wrapper-based algorithms tend to yield the same feature subsets, and the three filters provide relatively less similarity in general. Among all five feature selection methods examined, GainRatio and ChiSquared are determined as the best approach for our case study, because it identifies words relevant to the subject that highly correlate to a particular level (class) of PE even if they are sparsely represented in the dataset. We note that like most machine learning-based studies, the case study results are determined on the underlying dataset and the algorithms investigated. Our proposed approach, however, can be applied to other curiosity exercise datasets as well, and provide the relevant experts a better insight into the student data.

The rest of the paper is structured as follows: Section 2 details the case study methodologies including QFT, feature selection, modeling approach, and data preparation and processing; Section 3 presents and discusses the various results obtained from our case study; Section 4 concludes our paper with a brief summary of the work done and some directions of future work.

2. Methodology

2.1 Question Formulation Technique

A student's ability to formulate insightful questions is a critical life skill that enables the student to engage with the content for a deeper understanding and learning [21]. Questions serve the purpose of making clear and concrete that which is unknown or misunderstood by the student. By making the unknown concrete, a pathway for exploration, engagement and learning is opened to the student. As the student engages with the resources needed to answer the question, inevitably more questions are formed and new connections between topics are discovered. This process of questions driving deeper inquiry and learning is the premise of question-driven learning [1], sometimes referred to as inquiry-based learning [16].

Question-driven learning is hypothesized to stimulate curiosity and supports problem solving [1]. It has also been combined with Problem-Based Learning (PBL) in [17] to examine the role students' questions played in driving their learning. Students actively contribute to the development of a biology course through the questions they pose in [18]. Beatty et al. [19] present an audience response system combined with question-driven instruction to engage students in active knowledge building, as the instructor uses real-time formative feedback to tailor the classroom experience to student inquiry. An adaptive, question-driven intelligent tutoring system is developed and discussed in [20].

One framework that allows students to engage in question formulation as an exercise is the Question Formulation Technique (QFT) [11]. The QFT has been developed by the Right Question Institute [12] to empower students with the ability to formulate relevant and specific questions. The QFT involves a combination of (1) divergent thinking, (2) convergent thinking, and (3) metacognition, and is designed to be a collaborative exercise, ideally with groups of four students. One student should be selected as the recorder to record the generated questions.

The first stage of the QFT is called question-storming, in which the students generate as many questions as possible on a topic in a specified amount of time. The mode of thinking utilized during this stage is divergent thinking, as the students spontaneously form questions based on a prompt as soon as the question comes to mind. The prompt that introduces the topic to the students is called the question focus (or QFocus). The QFocus can be a statement, quote, set of images, video, audio clip, or any other type of prompt that sets the students on the path of generating questions on the desired topic. Typically, the QFocus is a provocative or outrageous statement, such as "Torture can be justified" [11]. Sometimes, it may be selected to emphasize a conceptual conflict, such as "For an RC circuit, forever is just five time constants away" [13]. There are four essential rules that govern the question-storming process to motivate the students stay on task of generating many questions while also encouraging a safe, inclusive space [11]. The first rule is to produce as many questions as possible in the allotted time. The second rule is to write the questions exactly as stated (including grammatical errors). The third rule is to not discuss or judge the quality of the questions during the question-storming process. The final rule is to try to formulate everything as a question. The instructor's primary function during the question-storming process is to encourage the students to adhere to the rules and cajole groups that are slow to generate questions. The instructor should not judge the quality of questions, neither with constructive feedback nor praise, as it undermines the divergent thinking process.

At the end of the question-storming process, the group should have many questions, some of which may be similar or complementary. The second stage is question refinement in which students eliminate equivalent questions, combine complementary questions to formulate multifaceted questions, eliminate grammatical errors, and generally improve the questions. This process involves convergent thinking, as students must analyze the questions to see how to improve the set of questions. The third stage is question prioritization. The instructor should provide some criterion or set of criteria on which to prioritize the questions. Some options include propensity for exploration, relevance to the topic, importance to the topic, question complexity, or level of student interest. The criteria selected by the instructor should be related to the desired purpose for which the questions will be used, e.g., a research paper, design project, or topic motivation [13].

2.2 Feature Selection Techniques

In machine learning, the typical task is to model a learner with the given dataset to predict a target feature (related to the given domain) based on a given number of predictor features. In the case of a dataset with a very large number of features and/or with the presence of data noise (especially feature noise), a feature selection (FS) process is performed prior to building the final predictive model. The former is applicable to our study where a token-based feature importance approach is taken, as explained in the Section 2.3 of this paper. We investigate five different FS approaches commonly used in the data mining domain, and they include [15]: two wrapper-based feature subset selection methods and three filter-based feature ranker techniques.

The wrapper-based approaches work by using a search algorithm (e.g., BestFirst and GreedyStepwise) to find a subset of features that collectively defines the performance of the classification model. A classifier is built using a given feature subset and evaluated using a performance metric. The classifier used in our study is the C4.5 decision tree, and the performance metric used is the Area Under the Receiver Operating Characteristic curve (AUROC). The ROC curve plots the true positive rate versus the false positive rate, for a given class. The wrappers yield a feature subset that collectively provide the best classification performance. Therefore, the size of the subset can vary for each dataset and no elements of the subset may be removed when building the classifier and presenting the results.

The filter-based feature ranker techniques consist of the ChiSquared, ReliefF, and GainRatio algorithms [15]. Rather than providing a subset of features as in the case of the wrapper approaches, the filters provide an ordered rank list of all the features from the best to the worst, based on a given performance metric. The ChiSquared attribute evaluator in Weka evaluates the worth of an attribute by computing the value of the chi-square statistic with respect to the class attribute. GainRatio is a modification of Information Gain by reducing its bias on highly branching features. It considers the number and size of branches when choosing a feature. This is done by normalizing information gain by the Intrinsic Information, which is defined as the information needed determine the branch to which an instance belong (the class label). The ReliefF algorithm computes a feature score by using the identification of feature value differences between nearest neighbor instance pairs. A “hit” occurs when a feature value difference is observed in a neighbor instance belonging to the same class, yielding a reduction in the feature score. Conversely, a “miss” occurs when a feature value difference is noted in a neighbor instance belonging to a different class, yielding an increase in the feature score. The distance function and the number of nearest neighbors is the key variants for ReliefF.

The open-source WEKA data mining and machine learning tool is used to implement our case study experiments, including the training of the classifiers and implementing the five feature selection algorithms [15]. In our study, all parameters other than the specific feature selection algorithms used and C4.5 classifier for the wrapper-based approaches in the Weka tool are set to default.

2.3 Modeling Methodology

2.3.1 Data Collection

A 400-level course in Artificial Intelligence (AI) was considered for our data collection purposes. The QFT methodology was applied to obtain the question-based responses to five QFocus statements (labeled in the form of, Qx), and they are:

Q1. AI did my homework. I did not cheat.
Q2. AI is the worst thing to happen to law enforcement.
Q3. AI has a singular moral code.
Q4. AI creates a more equitable job market.
Q5. AI algorithms should discriminate.

The data collection from the three stages of the QFT methodology were obtained from students of the course. To get a relatively decent size of dataset we focus our analysis and case study experiments on the questions obtained from stage two (question refinement) of the QFT methodology. To our knowledge there is no direct measure to evaluate a student’s curiosity degree, thus, we use an associated concept, Propensity of Exploration (PE), as a measure to provide insight into a student’s degree of curiosity. A panel of three domain experts (two faculty and a senior-class student) evaluated the stage two questions for their PE potential and scored them as either: Low (1) and High (2). The scoring of Low and High of the target feature PE, is used during the feature selection process for finding the tokens (words) in students’ questions that best reflect the different levels of PE, and thus, different degrees of curiosity. Thus, the PE scoring values are used as categories or classes to group the different question-based answers students developed in stage two.

2.3.2 Data Preparation and Processing

The question-based answers collected from participants was digitized into a Microsoft Word Document. We used the respective session number (each QFocus statement session) to label each document created. A final document was created in which we formatted the questions given by students by changing capital letters into lowercase letters and by removing any numbers, punctuation, and special characters. Subsequently, as mentioned earlier, after completing the data digitization process, each question is labeled as either Low or High according to their potential for Propensity of Exploration.

Every question in the dataset was analyzed to find the unique words it had. Toward this goal, we created a python script would open the Word documents using the docx Python library and find all the words in each question in the entire document. The QFocus prompt at the start of the document and the dashes to separate the prompt from the questions generated by the students are exempted from the word search. All the words were put into an array variable called "word_list" which was then looped through to find all the stop-words in the document. We used the stop-word list from the Natural Language Toolkit (NLTK) python package. This package can be installed with "pip install nltk." We compare every word in "word_list" to the words in the stop-word text file and added the non-stop words into another array variable called "filtered_words." This array variable still contained the PE ranking at the end of every question. To find all the unique words in the "filtered_words" array, we put every unique word into another array variable called "unique_words."

Lastly, we created two variables to calculate the number of unique words used per question and an array to write the occurrence of unique words used in a single question to a CSV file. We created a dictionary in which the unique word was set as the key, which defined the number of occurrences per question (called "num_words_dict"). A second variable, "num_words," was created with the intention of storing the occurrences of unique words from the "num_words_dict" in an array. To create and write into a CSV file, the CSV python library was used to insert the question number, occurrences of unique words in a question, and the PE label for the question for each QFocus session.

2.3.3 Modeling for Feature Selection

The WEKA data mining and machine learning too was used to conduct the feature selection experiments in our case study. As elaborated previously, the collected data was converted from text data into numerical form based on a text-to-tokens transformation and using a standard stop-words list. Each unique word that is not in the stop-words list is referred to as a token. Through feature selection, we obtained meaningful tokens (insightful for the PE metrics) and were able to reduce the sparsity of the high-dimensional sparse data set. The feature selection process was based on three filter-based feature rankers and two wrapper-based feature subset selection algorithms. The rankers included Chi-squared, GainRatio, and ReliefF, while the wrappers included BestFirst and GreedyStepwise search algorithms with the C4.5 decision tree as the classifier and AUROC as the performance evaluation metric.

3. Case Study Results

The frequencies of the Low-PE and High-PE question-based students text responses are shown in Figure 1, which shows the data for each of five QFocus statements, Q1, Q2, Q3, Q4, and Q5 (the QFocus statements are provided in Section 2.3.1). While there is not clear cut pattern across all the Qs, in general the High-PE questions are in higher numbers relative to the Low-PE questions. The exception being Q1, which could be reflective of students being new to the QFT process and in general responded with Low-PE questions.

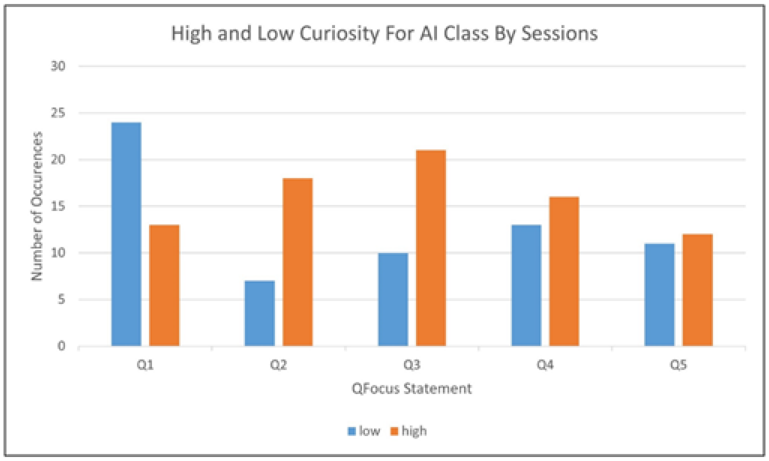

Figure 1: The PE (Curiosity Degree) Frequency of Low/High Questions

Best-First Search Based Wrapper FS				
Q1	**Q2**	**Q3**	**Q4**	**Q5**
use	ai	defines	used	discrimination
might	use	---	---	algorithms
build	---	---	---	good
homework	---	---	---	bias
GreedyStepwise Search Based Wrapper FS				
Q1	**Q2**	**Q3**	**Q4**	**Q4**
use	ai	defines	used	discrimination
might	use	---	---	algorithms
---	---	---	---	good

Table 1: Feature Subset Selection by the Wrappers.

Chi-Squared Ranker				
Q1	**Q2**	**Q3**	**Q4**	**Q5**
use	scenario	could	done	algorithms
might	ai	ai	remove	could
copyright	used	act	bids	people
cheating	maliciously	immoral	ai	ai
rules	times	way	specifically	begins
generated	breach	still	job	discriminate
considered	citizens	following	market	ethical
problem	privacy	code	certain	use
comes	ways	likely	sectors	oversees
person	people	scenario	biased	negative
Gain Ratio Ranker				
Q1	**Q2**	**Q3**	**Q4**	**Q5**
use	scenario	could	done	algorithms
might	ai	ai	remove	could
copyright	used	act	bids	people
cheating	maliciously	immoral	ai	ai
rules	times	way	specifically	begins
generated	breach	still	job	discriminate
considered	citizens	following	market	ethical
problem	privacy	code	certain	use
comes	ways	likely	sectors	oversees
person	people	scenario	biased	negative
ReliefF Ranker				
Q1	**Q2**	**Q3**	**Q4**	**Q5**
use	enforcement	define	might	discrimination
might	ai	gain	used	algorithms
past	worse	term	well	things
model	things	would	kinds	discriminate
less	cars	like	things	bias
programming	self	face	algorithms	thing
assignments	driving	follows	spell	good
give	scenario	scenario	different	avoid
teach	something	irobot	others	calls
study	hindering	us	make	train

Table 2: Features Selected by the Three Filter-based Rankers.

The feature subsets selected by the two search algorithms of the wrapper-based feature selection approach is shown in Table 1. The table shows the selected feature subset for each of the five QFocus statements, where the top half of the table represents the Best-First search algorithm's results while the bottom half of the table represents the GreedyStepwise search algorithm's results. Recall that a wrapper uses a classifier and a classification performance metric during its feature subset selection process. In our study we used the C4.5 decision tree classifier and the AUROC performance metric. Once a wrapper-based feature selection is done, any machine learner can be used with the selected feature subset to train and evaluate classifiers.

From Table 1, we observe that the two feature subset search algorithms generally yield similar or identical results. In the case of Q2, Q3, and Q4, the feature subsets are identical, and in the case of Q1 and Q5, the feature subsets are relatively similar. A deeper look at the features selected for each QFocus statement, we can notice interesting observations. For Q1 ("AI did my homework. I did not cheat."), the features are strongly reflective of the meaning of the statement in addition to extracting the "homework" token from the statement as an important feature. The tokens, "use" and "might," are semantically relevant to the Q1 statement, e.g., "I might have cheated." Similar observations can be interpreted from looking at the features selected for the other QFocus statements. For example, in the case of Q5, key tokens from the statement itself are observed as strong predictive features by the two wrapper-based feature selection methods. Finally, the tokens, "use" or "used", occur frequently in the table, which is intuitive given the different QFocus statements.

The feature selection results of the three filter-based rankers are shown in Table 2. After the word-to-vector tokenization process in our case study, the feature dimensionality was very large compared to the data point dimensionality. And since a filter-based ranker orders the different features from best to worst based on the predictive capability, we select the top 10 features to focus upon in this case study. This was done because in general the top 10 features, for a given ranker and QFocus statement, yielded the highest performance metric of the respective ranker. Among the features ranked by the three filters, as shown in Table 2, we observe that GainRatio or Chi-Squared provided identical token (for a given QFocus statement) both in features and their respective rankings. This was also observed in our previous study [21], where the QFT process was conducted for an Electric Circuits lower-level undergraduate course. The ReliefF ranker provided somewhat different top 10 features and their respective ranking for the different QFocus statements. With ReliefF, in general, the respective QFocus statements, both in terms of words and their semantics, yielded features that were intuitively or directly reflective of the respective statements. For example, with Q2, tokens such as "enforcement", "ai", and "worse" have a direct correlation with the QFocus statement, while tokens such as "hindering" and "scenario" provide a more semantic-based correlation with the QFocus statement.

4. Conclusion

The paper investigates data mining and machine learning techniques toward providing an insight into predicting the degree of curiosity a student has for a given course-related topic. To determine the level of curiosity among students engaging in an upper-level Artificial Intelligence course in an undergraduate computer science program, data is collected utilizing the Question Formation Technique. The latter collects text responses from students via a process with three stages, namely, divergent, convergent, and prioritization.

Data collection involves presenting students with thought-provoking five different QFocus statements, prompting them to formulate their responses in form of questions. The data is analyzed and interpreted through NLP for which Python-based scripts are developed toward an efficient organization of the student text responses, which is then analyzed using the WEKA data mining and machine learning tool. Feature selection is performed using three filter-based feature rankers and two wrapper-based feature subset algorithms. The course subject instructors determined that the extracted features provide meaningful insight into the "Propensity for Exploration" within the student text responses as a measure of their curiosity degree levels.

The five QFocus statements included: "AI did my homework. I did not cheat."; "AI is the worst thing to happen to law enforcement."; "AI has a singular moral code."; "AI creates a more equitable job market."; and "AI algorithms should discriminate." For the case study presented the best results were obtained with the Gain Ratio and ChiSquared filter-based rankers. While providing important features, the wrapper-based feature subset selection process yielded fewer tokens for the domain experts to analyze and evaluate the degree of curiosity (PE) in student text responses. Our future work will include performing the QFT and machine learning based approach presented here for other courses, both computer science courses and non-computer science courses.

References

[1] D. L. Schwartz, J. M. Tsang, and K. P. Blair. The ABC's of How We Learn: 26 Scientifically Proven Approaches, How They Work, and When to Use Them. W. W. Norton & Company, Inc., New York, NY, 2016.

[2] M. J. Kang, M. Hsu, I. M. Krajbich, G. Loewenstein, S. M. McClure, J. T.-Y. Wang, and C. F. Camerer, "The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory," *Psychological Science*, 20(8), pp. 963-973, 2009.

[3] T. G. Reio Jr. and A. Wiswell, "Field investigation of the relationship among adult curiosity, workplace learning, and job performance," *Human Resource Development Quarterly*, 11(1), pp. 5-30, 2000.

[4] J. Prusa, T. M. Khoshgoftaar and N. Seliya, “The Effect of Dataset Size on Training Tweet Sentiment Classifiers,” In *Proceedings of the 14th IEEE International Conference on Machine Learning and Applications* (ICMLA), pp. 96-102, Miami, FL, 2015.

[5] A Singh, J Ranjan, and M Mittal, “Big Data and Behavior Analytics”, Handbook of e-*Business Security*, Ch. 9, 2018. Taylor & Francis, CRC Group.

[6] P. Kendeou, P. Broek, A. Helder, and J. Karlsson J, “A Cognitive View of Reading Comprehension: Implications for Reading Difficulties,” *Learning Disabilities Research & Practice*, 29(1), pp. 10–16, 2014.

[7] L. McGarrity, “What Sentiment Analysis Can Do for Your Brand?” *Marketing Profs*, April 2016.

[8] A. E. Barth, S. Vaughn, P. Capin, E. Cho, S. Stillman-Spisak, L. Martinez, and H. Kincaid, “Effects of a Text-processing Comprehension Intervention on Struggling Middle School Readers,” *Topics in Language Disorders*, 36(4), pp. 368-389, 2016.

[9] E. M. Grossnickle, “Disentangling Curiosity: Dimensionality, Definitions, and Distinctions from Interest in Educational Contexts,” *Educational Psychology Review*, 28(1), pp. 23-60, 2016.

[10] T. J. van Schijndel, B. Jansen, and M. Raijmakers, “Do individual differences in children’s curiosity relate to their inquiry-based learning?” *International Journal of Science Education*, 40(9), pp. 996-1015, 2018.

[11] D. Rothstein and L. Santana. Make Just One Change: Teach Students to Ask Their Own Questions. Harvard Education Press, Cambridge, MA, 2015.

[12] The Right Question Institute, www.rightquestion.org.

[13] H. J. LeBlanc, K. Nepal, and G. S. Mowry, “Stimulating Curiosity and the Ability to Formulate Technical Questions in an Electric Circuits Course Using the Question Formulation Technique (QFT),” *IEEE Frontiers in Education Conference* (FIE). Indianapolis, IN. October 2017.

[14] G. Loewenstein, “The psychology of curiosity: A review and reinterpretation,” *Psychological Bulletin*, 116, pp. 75-98, 1994.

[15] E. Frank, M. A. Hall, and I. H. Witten, The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

[16] M. Pedaste, M. Mäeots, L. A. Siiman, T. De Jong, S.A. Van Riesen, E.T. Kamp, C.C. Manoli, Z.C. Zacharia, and E. Tsourlidaki, “Phases of inquiry-based learning: Definitions and the inquiry cycle,” *Educational Research Review*, 14, pp. 47-61, 2015.

[17] C. Chin and L.-G. Chia, “Problem-based learning: Using students' questions to drive knowledge construction,” *Science Education*, 88(5), pp. 707-727, 2004.

[18] M. Shodell, “The question-driven classroom: student questions as course curriculum in biology,” *The American Biology Teacher*, 57(5), pp.278-281, 1995.

[19] I. D. Beatty, W. J. Leonard, W. J. Gerace, and R. J. Dufresne, “Question driven instruction: teaching science (well) with an audience response system,” In *Audience response systems in higher education: applications and cases*, pp. 96-115. IGI Global, 2006.

[20] P. J. Muñoz-Merino, M.F. Molina, M. Muñoz-Organero, and C.D. Kloos, “An adaptive and innovative question-driven competition-based intelligent tutoring system for learning,” *Expert Systems with Applications*, 39(8), pp. 6932-6948, 2012.

[21] N. Seliya, H. LeBlanc, B. Hylton, Z. Youssfi, and M. Schweinefuss. Examining Trends in Curiosity Levels of Student Text-Based Questions. In *Proceedings of the 2019 American Society for Engineering Education Conference*, Tampa, FL, June 2019.