

Using Machine Learning Techniques to Assess
The Effect of a Focusing Statement on the
Quality of Question Generation

Reuben Thiessen
Stanford University
May 1, 2016

Abstract

In this study, we test if the addition of a focus statement to a question generation task can be correlated with questions that could be classified as questions that show evidence of utilization of higher-order thinking skills and also help a student to show divergent thinking in the creation of their questions.

As students become better question askers, they will grow in their critical thinking skills which will propel them to develop a deeper understanding of the material they are learning. Our research centers around one aspect of students becoming better question askers: could the addition of a focusing statement in a question generation assignment correlate with students creating questions that exhibit higher-order thinking skills (using Bloom's Taxonomy of the Cognitive Domain as a guide), and could students show divergent thinking as a result (using Guilford's Alternative Uses Task as a test for divergent thinking)?

Introduction

Asking questions is fundamental to learning. Not only is it core to learning in general, but it is crucial to being a critical thinker, which is an important skill in the 21st-century. According to The Right Question Institute, “the skill of question asking is far too rarely deliberately taught in school,” yet it is crucial part of students becoming lifelong learners.

There is unprecedented access to collective knowledge and information. It is unlike any other time in human history. However, Warren in “A More Beautiful Question,” submits that the “value of explicit information...is in what you can do with that knowledge, in pursuit of a query.” (Warren, p.23), There is good news! The act of asking excellent, creative questions is something that people can learn to do and improve with practice.

Research has shown that many college students lack the ability to access their understanding of academic content and material (Schwebel, Mahner and Fagley, 1990). Metacognitive strategies can help students understand content and use their time and attention more efficiently. Question-generation is an extremely valuable strategy that has been proven to increase metacognitive abilities and student’s comprehension and focus on main ideas on specific topics (Davey & McBride, 1986).

With the right support, question asking could become a reflex action when students are exploring new topics. But only if it is a skill that they have practiced and are using regularly. A student’s access to all the available information is only granted when the user can ask the right question. We believe that during this will help the student.

In a personal interview with the authors of the *Make Just One Change* book, they suggested that in fact it was the question focus that was essential to get students to ask better questions. In my personal experience working with students, we found that examples were important to helping students create more questions.

Thus our research question could be stated thusly: Will the introduction of a question focus in a question generation exercise make a difference in the creativity and divergent thinking shown in the generated questions?

General Description of Experiment

Our experiment involved two groups of students, the control group, which received no Question Focus statement, which we call the No Statement Group, and the treatment group which received a Question Focus in their instructions (the Question Focus group). We divided the students into two groups because we were testing just one variable: one group would see the Question Focus statement and one would not. Each group of students received an email asking them to visit a Google Form that included instructions to watch a short 5-minute video and then take approximately 15 minutes to type as many questions as they could think of into text boxes on the form.

We then took the questions generated by each student of each group and processed the question through a machine learning classifier to determine the confidence value for each question along each of the six levels of Bloom's Taxonomy, knowledge, comprehension, application, analysis, evaluation, and synthesis. After the questions were scored, we processed each user's group of questions through another machine classifier that scored along the four factors found in Guilford's Divergent Alternative Uses Test (the DT Test), which has been used to give a divergent thinking score. The four factors in the DT Test are originality, fluency, flexibility and, elaboration.

Our hypothesis was that the results would show that the introduction of a focusing statement in the question creation assignment would result in questions that could be categorized as higher-order thinking (scoring higher on analysis, evaluate and synthesize factors of Bloom's Taxonomy) and also have a higher score on the DT Test (the sum of the score of the four factors of the DT Test). If the Question Focus group scores higher on both the Bloom's Test and the DT Test, it would mean that the

addition of the question focus is a good addition to the process overall. Conversely, if the No Statement group scores higher on both, then the addition of the Question Focus would suggest that it is not beneficial to add the focusing statement. However, if the results are split, where one group scores higher on one metric and the other group scores higher on the second metric, then it would suggest that the addition of the question focus might be beneficial depending on the outcome desired by the teacher.

Methods

The experiment was delivered to two groups ($N = 5$) of students in a graduate-level class at an elite private university in the San Francisco bay area. The participating students were a convenience sample taken from same class as the study author. Each group consisted of three women and two men. One group was designated as the No Statement group and the other the Question Focus group.

Each group received an email asking them to visit a web page that contained a description of the activity, an embedded 5-minute video about the planet Mars and a web form that had spaces for 15 questions and a final space for any questions that the student wanted to fill out if they generated more than 15 questions. The Question Focus group's description included the sentence "The question focus statement may help to provide a launching point for your questions." Additionally there was a Question Focus statement that displayed the following sentence: Although humankind has learned much about Mars, there are still many mysteries about the planet. The Question Focus was also denoted by a bold title labeled "Question Focus Statement." There was no time limit on how long the user can take to create questions, but in the email the students received, it was suggested that the student spend at least 15 minutes generating questions.

Included in the instructions was a link to a reference card for creating questions that showed different examples of question types that were loosely aligned with Bloom's Taxonomy.

The responses were collected automatically by the web form in a file with each question delimited by commas, and each student's responses on a separate line. We decided to use a machine learning classifier to code the responses. After their responses were collected, the CSV file of the responses was processed by a Natural Language Classifier built on IBM Watson's Machine Learning platform, trained with questions tagged with the six levels of Bloom's Taxonomy (www.edupressinc.com):

- *Knowledge*: Exhibit memory of previously learned material by recalling facts, terms, basic concepts, and answers.
- *Comprehension*: Demonstrate understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating main ideas.
- *Application*: Solve problems to new situations by applying acquired knowledge, facts, techniques and rules in a different way.
- *Analysis*: Examine and break information into parts by identifying motives or causes. Make inferences and find evidence to support generalizations.
- *Evaluate*: Present and defend opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.
- *Synthesize*: Compile information together in a different way by combining elements in a new pattern or proposing alternative solutions.

Additionally a machine learning classifier written in the computer programming language Python was adapted from De Smedt's work "Modeling Creativity," based on Guilford's Alternative Uses Test (DT Test). The four factors of the DT are as follows:

- *Originality*: converts questions to numeric vectors and compares them using a cosine similarity formula, then clusters them. It scored it as follows: Questions that occur in 5% of the clusters get 1 point, questions that occur in 1% of the clusters get an additional point. Then they are divided by the total number of questions to account for the contamination problem (the test can be "gamed" by generating a large number of questions, dividing this number by the total number of questions normalizes the score.)
- *Fluency*: This is simply the total number of questions.

- *Flexibility*: This uses the Watson Natural Language Classifier trained on question categories (with a 86% precision-recall score) to determine the category of each question and count the number of discrete categories for each student.
- *Elaboration*: This factor uses a Python library to break each word in the question sentence into its part of speech, and then counts the number of sentences with more than two PNP (prepositional noun phrases) tags in each sentence. For example “under the table” or “as of today,” the idea is that a elaborate a sentence uses more prepositional noun phrases.

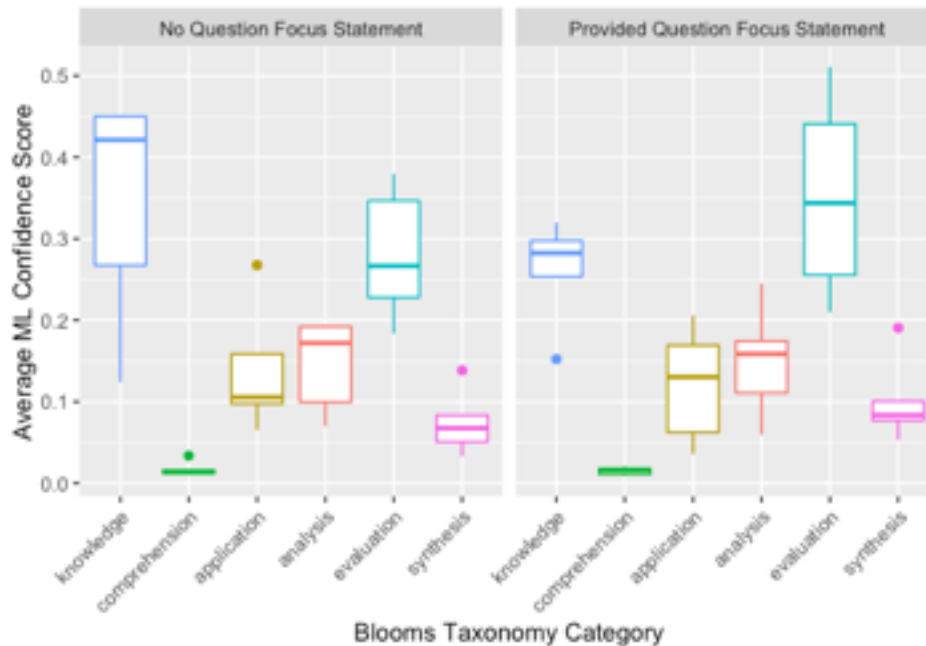
The machine classifier returned scores for each question and for each user. Then the results were averaged to compare the No Statement group with the Question Focus group.

Results

The Natural Language classifier returns the confidence score of the question sentence for each category of Bloom’s Taxonomy. These scores were then averaged to compare the two experiment groups. One can see from the table, that although both groups had a similar score for both application- and analysis-type questions, the No Statement group tended to have higher scores in knowledge and comprehension (lower-order) and the Question Focus group tended to have higher scores in evaluate- and synthesis-type questions. The classifier used had a precision-recall score of 0.75.

	Knowledge	Comprehension	Application	Analysis	Evaluation	Synthesis
Question Focus	0.26118133	0.014999806	0.120844842	0.149702216	0.352191489	0.101080316
No Statement	0.342796667	0.017265625	0.138946532	0.145397579	0.280933027	0.07466057

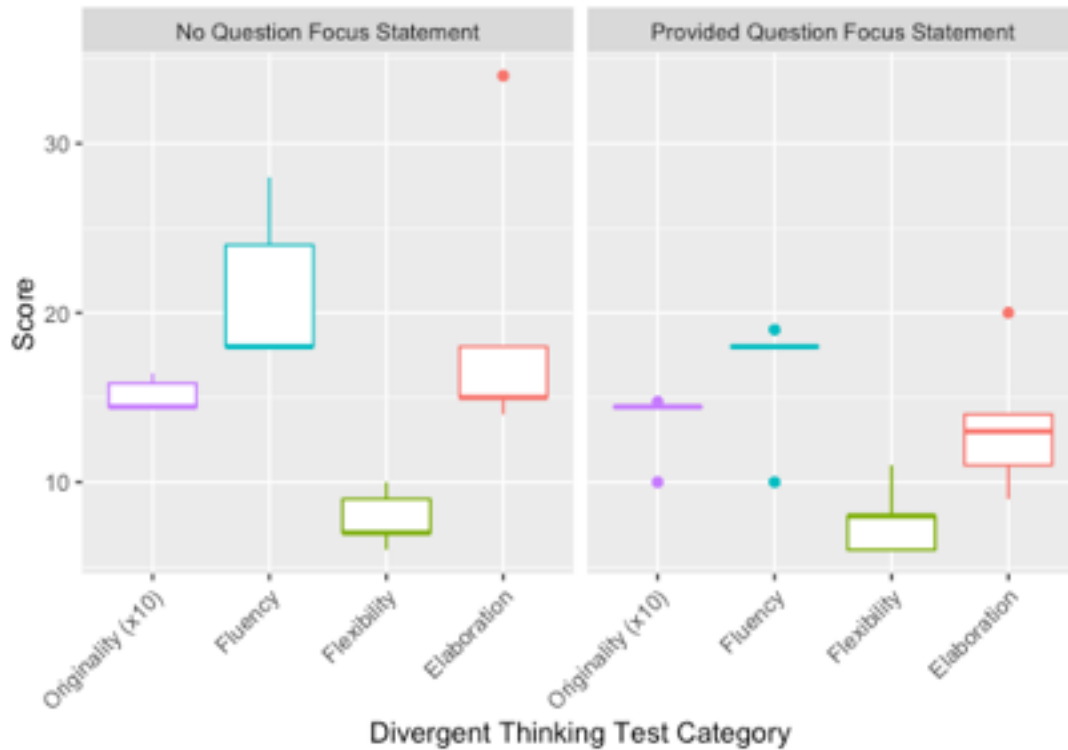
The box plots for all the questions for each group show in better detail that when provided with a question focus, the students tended to generate questions that were categorized higher on Bloom’s Taxonomy.



On the Divergent Thinking Test, the results were opposite: The No Statement group scored higher on the DT Test, also generating more questions (No Statement: 106 questions, Question Focus: 83 questions). The no statement group tended to have more elaboration in their questions and provide more original questions.

	Originality	Fluency	Flexibility	Elaboration	Total Average
Question Focus	1.361	16.6	7.8	13.4	39.161
No Statement	1.512	21.2	7.8	19.2	49.712

One can see that the No Statement group had more variation among the four factors. For the plot below, the originality score was multiplied by 10 for better visibility in the chart and to better show the variance between subjects



Discussion

The results show that our hypothesis was half correct. The Question Focus group showed the generated questions returned an average confidence score for Bloom's "higher-order" areas (analysis, evaluate, and synthesis). However, the No Statement group generated more questions, and scored higher overall on the DT Test.

After one sees the results as they are, it seems to be intuitive: if a student is given a focus statement, they will likely have less divergent questions, but still need to exhibit creativity around the questions they do create, if they are to generate a large number of questions. Conversely, a question focus statement might cause a student to not think in a divergent way as they have been predisposed to align their questions in the same topic area as the question focus.

The goal is to have students generate questions and develop question creation as a skill. There are potentially two general ways to start the question creation process: in response to a lesson, such as the experiment described here. The student receives some

sort of educational unit and then in response to the lesson, generates questions that were thought of because of the information presented. The second way is for the student to search and explore a new area of knowledge. This is where a student is given a topic, and then asked to explore and generate questions based on their own exploration.

Based on the results, we believe that a question focus would be helpful in the first scenario, to help the students come up with more creative questions for the material they just learned. The exploratory activity would benefit from more divergent thinking and thus a question focus statement may not be as helpful.

General Discussion

Question creation is an important skill to have, and has a direct impact on a student's learning. Adding a focal point for students when generating questions might be helpful when using divergent thinking is not as important as creativity around a specific topic area.

Some areas for future work include testing on a larger sample size. Additionally, not all students viewed the Reference card that had question examples, so ensuring that all students received and used the reference card may remove a lurking variable that the scores were impacted by students who viewed the quick reference card.

Our Bloom's question machine learning classifier had an accuracy score of 0.75. Increasing the accuracy of this classifier could help give greater validity to the results. Additionally, the use of a Google Form with 15 question blanks could have set an unintended suggestion for how many questions the students should generate and could have artificially inflated the number of questions created.

This study could be used as a framework for determining when to use a Question Focus statement in a question generation exercise. The teacher could determine whether the goal is using Bloom's level higher order thinking skills, or use divergent thinking to explore a new concept.

References

- Berger, Warren. A More Beautiful Question: The Power of Inquiry to Spark Breakthrough Ideas. Bloomsbury Publishing. 2014. Kindle Edition.
- Davey, Beth, and Susan Mcbride. "Effects of Question-generation Training on Reading Comprehension." *Journal of Educational Psychology* 78.4 (1986): 256-62. Web.
- De Smedt, T. (2013). *Modeling Creativity*. University Press Antwerp.
- IBM Watson Developer Cloud. (n.d.). Retrieved March 10, 2016, from <http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/nl-classifier.html>
- Indiana University. (n.d.). Creativity Test: Guilford's Alternative Uses Task (1967). Retrieved March 10, 2016, from <http://www.indiana.edu/~bobweb/Handout/d1.uses.htm>
- Rothstein, Dan, and Luz Santana. *Make Just One Change: Teach Students to Ask Their Own Questions*. Cambridge, MA: Harvard Education, 2011. Print.
- Schwebel, Milton, Charles A. Maher, and Nancy S. Fagley. *Promoting Cognitive Growth over the Life Span*. Hillsdale, NJ: L. Erlbaum, 1990. Print.
- Xin Li, Dan Roth, Learning Question Classifiers. COLING'02, Aug., 2002. <http://cogcomp.cs.illinois.edu/Data/QA/QC/>
- Bloom's Level Questions Training Data provided by:
Yahya, A. A., Toukal, Z., & Osman, A. (2012). Bloom's Taxonomy–Based Classification for Item Bank Questions Using Support Vector Machines. In *Modern Advances in Intelligent Systems and Tools* (Vol. 431, pp. 135–140). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-30732-4_17

Source code and training sets available:

<https://github.com/canuk/ml-paper-resources>