Regular and Automatic Feedback of Concepts as Formative Assessment

Samuel Cunningham-Nelson[1], Michelle Mukherjee[1], Andrea Goncher[2] and Wageeh Boles[1].
[1]Queensland University of Technology, [2]Charles Sturt University
Corresponding Author Email: samuel.cunninghamnelsonn@qut.edu.au

STRUCTURED ABSTRACT

CONTEXT

Regular formative assessment shows both the development and the gaps in student knowledge and understanding as they progress through a unit of study. The formative learning cycle is effective when students can receive feedback in real-time and address any misconceptions. Online platforms provide good mechanisms for feedback; however, the current systems are often limited to multiple choice questions or keyword detection in short response tests, and typically cannot assess in-depth conceptual understanding.

PURPOSE

In this study, we explore the use of machine learning techniques and an online interface to provide constructive feedback to students regarding their conceptual development in electrical engineering. This will be exploited in the context of the following research question:

- How can we use text analysis to provide real-time feedback to students' responses to improve their learning?

APPROACH

A system is presented which utilises an easy to use online interface that allows students to enter their responses to questions in a set of quizzes that test their conceptual understanding. Using various textual analysis techniques, the system can provide students with immediate feedback on both short response text and multiple-choice questions. The system is designed to provide feedback on more conceptual, and deep learning type questions.

RESULTS

The developed system aimed to provide electrical engineering students with immediate feedback on formative assessment tasks. We surveyed several second-year students who used this online system, to gauge their interest and opinions, with positive results. Further, to enable lecturers to track students' performance, and take appropriate follow-up actions, lecturers can see a summary overview of how the students performed in answering the quizzes. The uptake and usage in the chosen subject are discussed and presented.

CONCLUSIONS

By using the proposed system of automated formative testing and feedback, we conjecture that it would be possible to identify the gaps in conceptual development of a larger number of students in a shorter timeframe than previously possible. This will also provide students with opportunities to assess their understanding, and thus develop self-efficacy in learning. Also, lecturers would be able to monitor students' responses, and take appropriate actions to enhance student learning. Analysing these results will help to better respond to the diversity in learners and give the educator some further insight into possible teaching actions and activities that can be taken.

KEYWORDS

Conceptual Understanding, Text Analysis, Student Feedback

Introduction

Overview of Formative and Realtime Feedback

Assessing students and providing meaningful feedback is a difficult task (Ferguson, 2009). While it is time-consuming for teaching staff to provide regular, detailed feedback, many studies show that feedback is often not utilised (Race, 2010; Cathcart, Greer, & Neale, 2013). Even if an educator spends time to develop informative feedback this may go unread or arrive at an ineffective time. This highlights the need for automated feedback, which can be given in a timely manner.

Educators may have limited knowledge of the diversity of learners in higher education classrooms due to large class sizes and lack of formative assessments. Students that enter a course can have diverse prior knowledge, experiences, and backgrounds. Introducing formative assessments accompanied by feedback is one way to gain insights into the range of students' prior knowledge, as well as how they are progressing.

Multiple Choice Vs Short Response Questions

Multiple-choice questions are an attractive option for educators because they are easy to mark. The disadvantage of these multiple-choice questions, however, is that students can guess the answer, or they may not have a complete understanding of the concepts being tested in the question. To combat this, a free-text response component can be added (Goncher & Boles, 2017; Cunningham-Nelson, Goncher, Mukherjee, & Boles, 2017). This free-text component allows students to justify their chosen answer. This gives two benefits: (1) it allows a reader to check if a student has a proper understanding of the concept being tested and (2) allows the student to think more carefully about the answer they are selecting, and why. However, adding this free-text component currently introduces a disadvantage; we lose the ability to automatically mark the multiple-choice questions.

Text Analysis

What if we can automatically mark these text responses? This would allow a user to keep the benefits of a fast feedback response, as well as the benefits of free-text comments. The field of text analysis provides some insight into this. One of the main applications of text analysis, is sentiment analysis - understanding the positivity or negativity of a statement. Sentiment analysis is commonly performed on pieces of text such as twitter posts (Pak & Paroubek, 2010), product reviews, and movie reviews (Jong, 2011). With the growing amount of online text data, this is becoming increasingly useful.

Research Question

Combing the use of multiple-choice, free-text responses, and ensuring that the feedback is automated, we endeavour to answer the following research question:

- How can we use text analysis to provide real-time feedback to students' responses to improve their learning?

Method

CCU Format

The "Check Conceptual Understanding", or CCUs, are small conceptual style quizzes – different to traditional quizzes. The CCUs were released on a fortnightly basis, one for each topic in the unit. They were designed for students to complete after each main content area was presented and discussed in class, before moving onto the next topic. They were

designed to be formative, and with fast completion in mind to ensure that they were not seen as an extra load or burden for students.

Each CCU consists of three questions. The first question presented is from the Signals and Systems Concept Inventory (Wage, Buck, Wright, & Welch, 2005). Concept inventory questions have been carefully designed by experts targeting areas in which students might have misunderstandings. The options presented, along with the correct one includes several distractors aligning to common misunderstandings students may have about that topic. This conceptual question then has a free-text component added, asking students to justify their multiple-choice selection. Examples of these questions are shown in previous work (Cunningham-Nelson, Goncher, Mukherjee, & Boles, 2017). Two multiple-choice only questions then follow this, based on related content areas.

Text Analysis

With the free-text response component added to the CCUs, the textual part of the response needs to be automatically marked and classified. An initial bank, consisting of a previous 100 responses were graded by a tutor to get an initial benchmark. Previously discussed in further detail (Cunningham-Nelson, Goncher, Mukherjee, & Boles, 2017), the textual responses were pre-processed, modelled as word vectors (Mikolov, Chen, Corrado, & Dean, 2013) and classified using a Gaussian Support Vector Machine classifier (Joachims, 1998).

Graphical Interface

The graphical interface is designed with ease of use in mind. It was also important that students could get instantaneous feedback. The CCUs were offered in the learning management system for students to access periodically throughout the semester. The interface allowed students to quickly see their errors, or correct answers.

The interface allows students to answer multiple choice questions, and then justify their response with an added textual component. Feedback is then be presented to students on the same interface underneath each answered question.

Survey

A survey was constructed to assess students' perceptions of the CCUs and allow students to provide additional feedback. The questions used were validated (Evergreen, Gullickson, Mann, & Welch, 2011); given to several colleagues, as well as three students from the class. Following the validation process, the questions were narrowed down. These questions can be seen in the Appendix.

This survey was run during the final tutorial sessions of the semester as an optional questionnaire for students who attended this tutorial to complete. It was anonymous, with no student information collected.

Results and Discussion

We examine usage, performance, and accuracy of the CCUs implemented in their initial pilot. Each of these represents an important component in verifying the legitimacy of the CCUs.

Usage

First, we will examine the usage of the CCUs throughout the semester.

Table 1 shows three columns; the number of times students accessed to the CCU link, the number of time students submitted their answers to the CCU, and the percentage of student click-throughs which were submitted. From this table we can see an initial higher level of interest in click-through for the earlier CCUs, but a small percentage of those students actually submitted.

Table 1 - CCU Usage

CCU number	Number of click-throughs	Number of submissions	Percentage submissions from click-throughs
1	354	137	38.7%
2	198	98	49.5%
3	127	83	65.4%
4	111	79	71.2%
5	96	53	55.2%
6	62	53	85.5%

Breaking the usage down further into dates throughout the semester, Figure 1 shows the number of daily CCU submissions. Aligning with other research looking at worked example videos, exam preparation is often a driving factor for students using resources (Barns, Pickering, & Dawes, 2017). Figure 1 confirms that some students use the resource throughout the semester, however the vast number of submissions occurred in the week leading up to the final exam. It is also important to note that looking at the submissions logs, students often attempted the CCUs multiple times to check if their reasoning and answers were correct. These statistics and usage numbers demonstrate that students found the CCUs a useful tool when studying and preparing for their final exam.

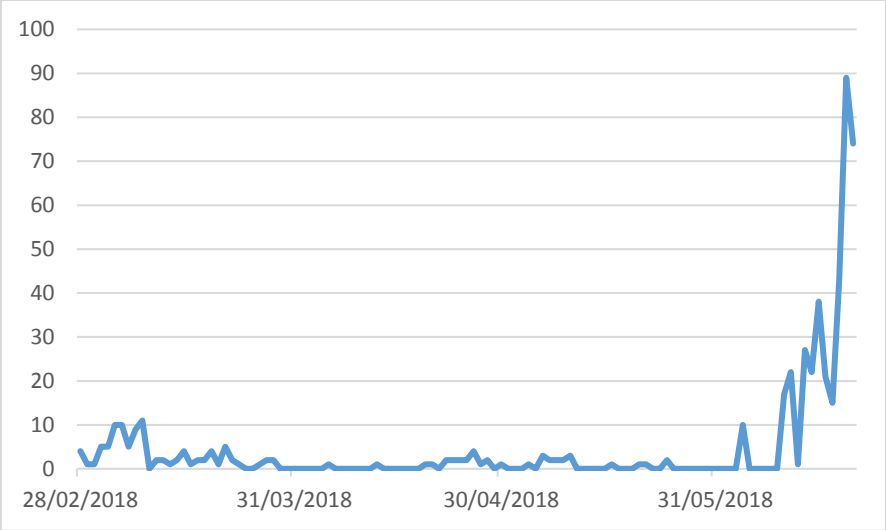

Figure 1 - Number of Daily CCU Submissions

Perception

Student perception of the CCUs was an important aspect to measure throughout this initial pilot run. We surveyed **28 students** in the unit. Although the sample is not large enough to provide an absolute proof, it does give an initial benchmark and insight. The six survey questions presented can be grouped into three main areas: interface, feedback and promotion.

Figure 2 shows the students' feedback based on the visual interface used. The majority of students were satisfied, or very satisfied with the interface presented, and no students were dissatisfied with it. This is a positive note, and some further suggestions were made from students on the placement of certain elements on the interface, and size of images presented in the questions. These can be incorporated in the next iteration.

Figure 3 shows the students' comments in their responses to the survey questions based on the feedback given to students from the CCUs. Similarly, to the survey comments on the

user interface, students' opinions were positive overall about the CCU feedback presented – the majority of students indicated that they were satisfied or very satisfied.

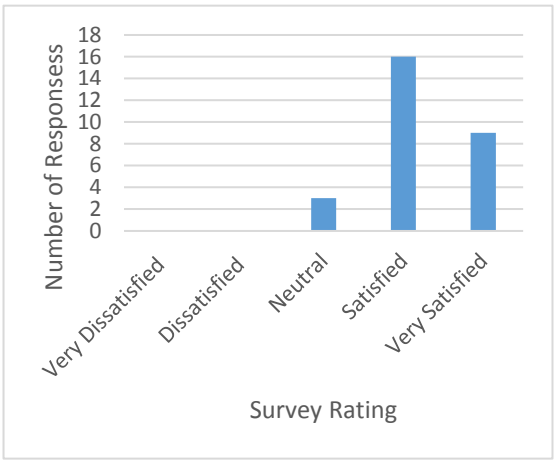

Figure 2 - Interface Feedback

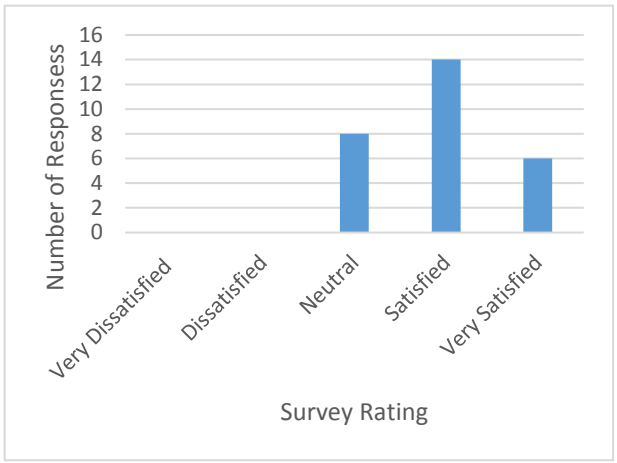

Figure 3 - Feedback Satisfaction Data (Q4)

Figure 4 shows a summary of responses to the multiple selections of question 5 in the survey. In this question, students were asked to mark all options that apply out of the four options presented. The orange line in

Figure 4 represents the total number of students surveyed (28). Twenty-two out of twenty-eight students agreed that the CCUs helped to identify gaps in understanding. This is extremely positive as allowing students to self-identify gaps in understanding was one of the main goals of the CCUs, and is important for students to know these gaps, and for educators to know where do major gaps exist.

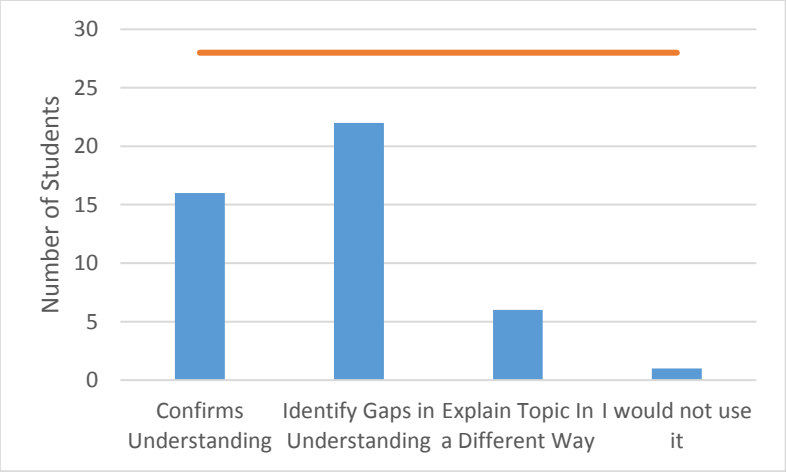

Figure 4 - Question 5 Perception

In addition to the numerical analysis above, students were given the option to add a free-text response and provide additional comments on the experience. Several positive comments were given about these CCUs including: *"It was well done"* and *"Good tool, should be used in all units"*. Constructive feedback was given from students and helps to identify some areas for future improvement. These comments include: *"More questions asked to ensure they fully understand"*, *"More detail on why you are wrong. Link to relevant lecture / tute material"* and *"Make the questions worth a percentage of the grade"*.

An interesting comment from a student identifies a possible disconnect for the CCUs presented: *"I feel there is a huge difference in these questions compared to the expected understanding for assessment. Making these pretty useless"*. This highlights the need in the

future to emphasie the link of these conceptual style questions to the numerical ones presented in the tutorials, so they do not appear to be separate.

Accuracy

Since students are being provided with instantaneous feedback based on their text responses, it is important that their validity and accuracy be verified. The textual responses that students submitted were automatically classified (eGraded) as correct or incorrect. At the end of the semester, the previously submitted student responses were then manually marked by a tutor, to test the validity.

As presented in Table 2, we can see the accuracy for each of the 6 CCUs. The first two rows are where the responses should ideally sit, where the manual marking of the student responses agrees with the automatically eGraded classification. The 3rd and 4th rows show a disagreement between the manual and eGrade mark. The value of N represented the total number of submissions for that CCU. The False positive rate is the percentage of responses labelled correct, when they should be incorrect. The False negative rate is the percentage of responses labelled incorrect, when they should be correct. The final correctly classified rate is the overall percentage of responses where the manual mark agreed with the automated mark. Looking at CCU2 as an example, an overall accuracy of 97% was obtained. Only 3 out of 98 text responses were classified incorrectly, showing very promising results.

Table 2 - CCU Accuracies

	CCU1	CCU2	CCU3	CCU4	CCU5	CCU6
Student correct, eGrade correct	104	22	48	28	15	35
Student incorrect, eGrade incorrect	19	73	24	16	34	10
Student correct, eGrade incorrect	10	3	2	6	4	4
Student incorrect, eGrade correct	4	0	9	29	0	4
N (Total)	137	98	83	79	53	53
False Positive	2.92%	0.00%	10.84%	36.71%	0.00%	7.55%
False Negative	7.30%	3.06%	2.41%	7.59%	7.55%	7.55%
Correctly Classified	89.78%	96.94%	86.75%	55.70%	92.45%	84.91%

The performance for most of the CCUs is not very different from the variance expected in human marks (Suto, Nádas, & Bell, 2011). CCU5 however provided some concerning results, with a low accuracy of 56%. Spelling errors were one contributing factor for incorrectly labelled responses. This warrants further investigation.

Conclusions and Recommendations

Feedback is an extremely important part in students' learning experiences. Without proper feedback, students will often struggle, and misconceptions linger in students understanding. The format for CCUs presented helps to address misconceptions and identify the range of a student cohort's understanding.

Results from the pilot semester show high student satisfaction, a good usage, and promising accuracies for the textual classification. Future work includes investigating various pre-processing and classification techniques to improve the accuracy of the system. We also plan to investigate further ways in which overall feedback can be presented and implemented in class.

The CCUs extend beyond traditional multiple-choice quizzes. The added textual component means that students need to think more carefully before selecting their answer and must justify their response. The ability to automatically provide feedback, allows students to receive it in-time, while they are thinking about the topic, and helps them to self-identify gaps in understanding. In addition to this, the benefits for the educator include minimised time spent on individually identifying student misconceptions, and the capability of identifying the areas which the entire class is excelling or struggling with. Formative feedback approaches, such as the CCUs, can address diversity in learning and improve student engagement through timely feedback. Once setup, these CCUs are able to operate independently from the educator.

References

Barns, S., Pickering, E., & Dawes, L. (2017). Worked example videos as a valuable blended learning resource in undergraduate engineering units. *28th Australasian Association for Engineering Education Conference (AAEE2017).* Manly, Sydney, NSW.

Cathcart, A., Greer, D., & Neale, L. (2013). Learner-focused evaluation cycles: facilitating learning using feedforward, concurrent and feedback evaluation. *Assessment & Evaluation in Higher Education*, 790–802.

Cunningham-Nelson, S., Goncher, A., Mukherjee, M., & Boles, W. (2017). Pointers to Conceptual Understanding. *Proceedings of the 28th Australasian Association for Engineering Education (AAEE).* Manly, Sydney, Australia.

Evergreen, S., Gullickson, A., Mann, C., & Welch, W. (2011). Developing and Validating Survey Instruments. *Web seminar Evaluate. www. evalu-ate. org. 269/387.5895. Western Michigan University*.

Ferguson, P. (2009). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education, 36*(1), 51-62.

Goncher, A. M., & Boles, W. (2017). Enhancing the effectiveness of concept inventories using textual analysis: investigations in an electrical engineering subject. *European Journal of Engineering Education*, 1-12.

Hamel, F. L. (2003). Teacher understanding of student understanding: Revising the gap between teacher conceptions and students' ways with literature. *Research in the Teaching of English*, 49-84.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *European conference on machine learning* (pp. 137-142). Springer.

Jong, J. (2011). Predicting Rating with Sentiment Analysis. *Stanford Univ.*

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Nips*, 1-9.

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *LREc*.

Race, P. (2010). Making Learning Happen.

Suto, I., Nádas, R., & Bell, J. (2011). Who should mark what? A study of factors affecting marking accuracy in a biology examination. *Research Papers in Education, 26*(1), 21-51.

Wage, K. E., Buck, J. R., Wright, C. H., & Welch, T. B. (2005). The Signals and Systems Concept Inventory. *IEEE Transactions on Education, 48*(3), 448-461.

Acknowledgements

We would firstly like to acknowledge the authors of the Signals and Systems Concept Inventory for allow us to use it in this study, and the students who provided responses to the questions. The authors would also like to thank Dr Dhammika Jayalath for his efforts in implementing this concept inventory over several years. Finally, they would like to thank Justin Kennedy for his help in testing and validating the CCUs.

The study was approved (1600000964) by the QUT human research ethics committee.

Appendix

1. How satisfied are you with the user interface for presenting and entering information?
2. Do you have any suggestions to improve the graphical interface?
3. How satisfied are you with the feedback given for your answers to the questions – specifically the feedback on your text justification for Question 1?
4. Do you have any suggestions to improve the feedback given?
5. What would you use this for? (Please select all those that apply)
    a. Using this tool confirms to me that I understand the topic.
    b. Using this tool helps me to identify gaps in my understanding.
    c. This tool has helped me to learn by explaining the topic in a different way.
    d. I would not use it – this tool does not help me to learn the topic.
6. Do you have any suggestions on how to promote this, and encourage students to use it?