A three-year longitudinal textual analysis investigation of students' conceptual understanding: Lessons learnt and implications for teaching

Samuel Cunningham-Nelson[a], Andrea Goncher[b] and Wageeh Boles[a]
Queensland University of Technology[a], Charles Sturt University[b]
Corresponding Author Email: samuel.cunninghamnelson@qut.edu.au

CONTEXT
Concept inventories are specific tests, designed to elicit misunderstandings or misconceptions. They are a set of multiple-choice questions (MCQs), designed to include the correct option, as well several distractors (Libarkin, 2008). One attractive advantage of the ease of scoring concept inventories, is their MCQ format. However, this same format does not detect whether students arrived at their correct answers by pure guessing. By adding a space for students to add a textual justification (Goncher, Jayalath and Boles, 2016), their answers can be checked to ensure that the concepts are correctly understood.

PURPOSE
The purpose of this research is to address the following research questions:
1. What are the similarities and differences observed in students' conceptual understanding scores, and persistent misconceptions?
2. How can textual analysis help discover the links between students' scores on conceptual understanding tests, their confidence, and possible teaching strategies?

APPROACH
There are 2 Signals and Systems Concepts Inventories, one for the continuous domain and one for the discrete (Wage, 2005). These two exams were run at both the beginning, and the end of semester, over a period of three years. Data collected from these contains a multiple choice response, and textual explanation for each for the 15 questions used in this study. This data was analysed using Leximancer (textual analysis program) and MATLAB to extract concepts from these responses, and links between the concepts and students' misconceptions were able to be identified.

RESULTS
Comparing the tests taken over several years, trends were analysed and compared. Since this has been run over multiple years and some actions were taken to address some problem areas and continuously occurring problem areas were identified. Rates of guessing in particular topic areas were also analysed and examined.

CONCLUSIONS
From this study, the concepts causing persistent difficulties to students were explored. Several persistent misconceptions were identified over the 3 years, both with low correct rates, and high guessing rates. The added textual component to the usual multiple choice response made this inferences possible.

KEYWORDS
Concept Inventories, Textual Analysis, Conceptual Understanding, Misconceptions.

Introduction

Engineering education efforts that focus on improving student learning recommend moving away from only assessing and teaching quantitative skills to developing a strong conceptual knowledge base. Facilitating students in developing conceptual knowledge requires meaningful assessments that reveal student thinking (Pellegrino, 2001). The need for innovative assessments that utilise students' written responses is important to providing instructors (and students) insight into their reasoning.

Formative assessments can provide insights into student understanding at different intervals of their instruction and learning. Open and constructed response questions require students to present their written, or text-based, answers, and can provide more information about their understanding than standard multiple choice assessments (Birenbaum, 1987; Popping, 2012). In order to increase the capability to formatively and accurately assess students' conceptual understanding, our study focuses on automating the collection and analysis of students' written responses. Our approach uses students' textual responses, and applies text analysis techniques to process the information and measure students' conceptual understanding.

Assessing Conceptual Understanding

Well-designed multiple-choice tests use questions and possible answers, i.e. distracters, that are designed based on research about how students understand the concepts in a specific area. Concept inventories are specific assessments, designed to elicit misunderstandings or misconceptions. They are a set of multiple-choice questions (MCQs), designed to include the correct option, as well several distractors (Libarkin, 2008). One attractive advantage of the MCQ format is the ease of scoring. The drawbacks of concept inventories are that they need to be designed by an expert (Arbogast, 2016), and usually are designed to test specific concepts within an identified domain, e.g. signal processing. The multiple choice nature of concept inventories introduces a validity threat by requiring students to select one of four or five possibilities, imposing a "forced choice" (Prevost et al., 2013). Assessments that capture a more explanatory answer can reveal how the student arrived at the selected choice.

The MCQ format also does not detect whether students arrived at their correct answers by a process of elimination or pure guessing. By adding a form field for students to add a textual justification (Goncher, Jayalath and Boles, 2016), their answers can be analysed to reveal if the selected MCQ response represents concepts are correctly understood. The findings from students' explanatory answers can be used in the development and teaching interventions, which are based on the learners' alternative conceptions. Given the importance of conceptual understanding in engineering, we wanted to examine if students accurately understand the concept associated with each correct selection, or if the answer was selected from guessing or a process of elimination. Furthermore, if a student does not answer correctly, how can we find out what types of misconceptions they have regarding a particular conceptual area. In this paper, we analysed the student responses to concept inventory questions, over a three-year period, and outline how MCQ and text responses might be combined with each other to provide meaningful insights into students' conceptual understanding. This work extends (Goncher, Jayalath and Boles, 2016) by adding an extra year of conceptual inventory data, and further interpretation, focussing on guessed questions.

Text Analysis Tools

The methods and software applied in our study involve natural language and text analysis algorithms. We aim to automate the extraction of key concepts and their relationships to one another by using available domain and semantic knowledge from commercially developed software. We also intend to build on, and suggest, additional categories for improvement.

Our researching in the signal processing domain can improve the ability of the software to learn from the manual assignments, especially when the software is not able to automatically assign concepts to categories in the signal processing domain.

Leximancer

Leximancer is a textual analysis program that can analyse and produce visualisations of the concepts and concept clustering within a text document. Leximancer is able to quantify and explore relationships between words and concepts for large text documents based on a classification system of learned lexical concepts. The algorithms are based on the Bayesian decision theory approach to prediction. Similar to a two-way contingency statistic, Leximancer accounts for how frequently two words occur together and how frequently they occur apart. This metric represents a relationship between concepts based on given fragmented information. We use Leximancer to investigate how the concepts contained in students' textual responses were connected to one another to see if the concepts were related in complex or simplistic ways.

MATLAB

We used the string compare function of MATLAB to analyse the text responses that indicated a "guessed" response. A string is a character array, and the string compare function allowed us to search the body of text responses for cases that indicated some level of uncertainty. We assigned strings to include 'guess', 'guessing', 'uncertain', 'not sure', and 'no clue'.

Research Questions

We argue that an assessment instrument that incorporates a written component, capable of being analysed automatically gives instructors the opportunity to adjust their teaching to eliminate or at least reduce misconceptions. The categorized textual explanations may be used to supplement or validate the statistical information gained from the MCQ selections. The goal of this study was to identify the types and categorizations of misconceptions not uncovered by the distractor selections, through text analysis of student explanations. This study extends the application of conventional concept inventories and other measurement-based applications used to evaluate conceptual understanding. We structure our analysis by investigating the following research questions:

1. What are the similarities and differences observed in students' conceptual understanding scores, and persistent misconceptions?
2. How can textual analysis help discover the links between students' scores on conceptual understanding tests, their confidence, and possible teaching strategies?

Method

Research Design

We employed a multi-year, case study in an electrical engineering context, focusing on signals and systems material. Using short answer, text responses, our process incorporated a component for an online signals and systems concept inventory test to capture written explanations of students' selections. The format required students to select one of the multiple-choice options and write an explanation for their selection in short answer format.

Instrument: Signals and Systems Concept Inventory

The Signals and Systems Concept Inventory (SSCI) is a 25-question multiple-choice exam developed to assess core concepts in undergraduate signals and systems courses. The continuous and discrete time versions of the SSCI was created and validated to reveal any student misconceptions by developing multiple-choice questions that assess certain concepts and includes distractors that represent common misconceptions [Wage, Buck, Wright, Welch, 2005]. Like other concept-based Concept Inventories(CIs) the SSCI is visual in nature and does not require much mathematical computation and tests mathematical knowledge through graphical representations.

We used a subset (15 questions) from the complete discrete and continuous SSCIs to allow students to provide written answers within a reasonable testing time. The number of concept inventory questions was reduced to allow us to focus on the areas of interest for the relevant unit, as well as reducing the time and pressure on the students. The selected questions covered five fundamental conceptual areas in signals and systems. Several of the questions were selected because they are identified as linked or synthesis questions, or questions that test concepts required to understand the synthesis question. The study described in this paper was carried out over 2014, 2015, and 2016 in three sections of a fourth year unit in electrical engineering at an Australian university. This was a large-enrolment course with sections of 96, 73, and 85 students, for the respective years 2014-2016.

Data Set

The Signals and Systems Concept Inventory was designed for two versions: Discrete and Continuous. These two concept inventories tests have been administered in a fourth year unit at the Queensland University of Technology. Both versions were administered as the discrete concept inventory tested concepts (such as sampling) that were not tested in the continuous concept inventory. This unit, focussed on digital communication concepts is the third in a series of signals analysis and telecommunication units. The Signals and Systems Concept Inventory was run in both the first and final weeks of semester – as a prior and post test. The prior test at the beginning allows the initial understanding of that cohort to be gathered, and then content delivery modified in response to insights gained from students' answers. The post test then evaluates the same questions, to test for an improvement throughout the duration of the subject.

Run in 2014, 2015 and 2016 data has been gathered from various students, and 189 entries between the concept inventories over three years have been completed.

Concept and Feature Extraction

The software used, "Leximancer" allows the links between words and concepts to be viewed in a visual manner. Themes or concepts are identified and grouped together within Leximancer, which allows any duplicate concepts to be captured. Irrelevant or unimportant themes can also be removed from this process.

Data Pre-processing

Data pre-processing, or data conditioning is an important part of the analysis process. It involves getting the data prepared, and ready to work with. For this work, the data pre-processing involved grouping responses into correct and incorrect categories, based on the corresponding multiple-choice selection for each question, as well as breaking each text response into individual words so that these could be examined one by one.

Results

Multiple Choice

Initial results for the three years in which the Signals and Systems Concept Inventory was administered is displayed in Table 1. The table includes, for each year, the two types of tests, as well as the prior (beginning of semester) and post (end of semester) data. For each test three key pieces of information are given: μ (average or mean), σ^2 (standard deviation) and

N (sample size or population). We did not administer the continuous version of the Signals and Systems Concept Inventory in 2014, so results are not included in the table.

Table 1: Detailed Summary of Results Year by Year

	Discrete						Continuous					
	Prior			Post			Prior			Post		
	μ	σ²	N	μ	σ²	N	μ	σ²	N	μ	σ²	N
2014	9.21	2.36	61	9.93	2.40	14	-	-	-	-	-	-
2015	9.77	2.41	22	11.40	2.46	10	10.41	2.40	22	10.56	2.74	9
2016	8.29	2.44	17	8.70	2.16	10	8.24	2.46	17	11.41	2.07	7

These pieces of information provide a valuable starting insight into the progress of a given cohort. For example, from these results we can make conclusions such as:

- Over each year and each test type, an improvement in the average score is seen.
- There is no noticeable trends or improvements between the three cohorts.
- The largest mean value is the 2016 continuous post CI, a value of 11.41.

The sample size for the concept inventories above is smaller than the number of students enrolled in the corresponding digital communications course as the CI was optional for students to take part in. The number of students taking the post test is also significantly lower than those taking the pre test. This is most likely due to the timing of the post test, as it is run at the end of the university semester.

The results above demonstrate an overview for the test as a whole, however examining the results grouped into topics can give even more insights into the performance of a cohort. It is important to take these results as one instance, and not a general result. Some sample sizes, especially for the post tests, are not large enough to warrant a general assumption.

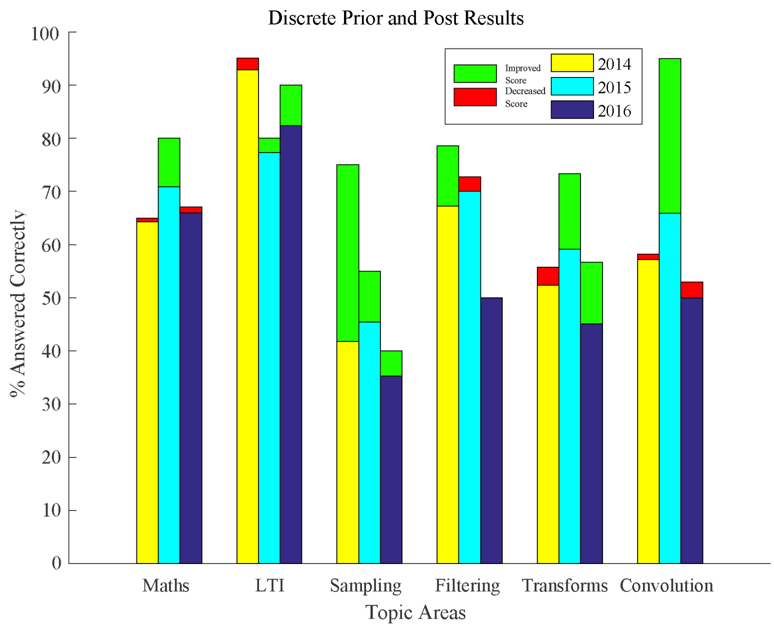

Figure 1: Discrete Test Results

Examining the results graphically, Figure 1 and Figure 2 below show results for the discrete and continuous concept inventories. The topic areas (for the discrete test) identified were: Maths, Linear Time Invariant (LTI) systems, Sampling, Filtering, Transforms and Convolution. This allows particular topic areas or concepts which students might be lacking in to be easily identified and focussed on in class. Each topic area has results for each year: 2014, 2015 and 2016. The figures identify also, marked in red and green, the amount of change between the prior and post tests. Green indicates an improvement in scores, and red a decline.

As an example, looking at the sampling topic area in the discrete concept inventory, the percentage of questions answered correctly was lowest out of all the tested sections. This also corresponds with significant improvements in the same area between the prior and post tests.

Figure 2 shows the continuous concept inventory which was implemented over 2015 and 2016.

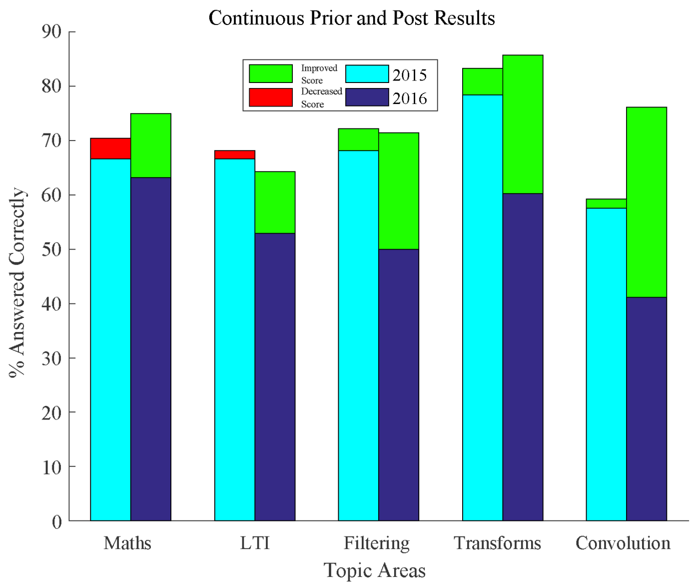

Figure 2: Continuous Test Results

Text Analysis

The text response component provides valuable insight into the understanding of a given concept. Students completing the test were asked to use words such as “guess” or “elimination” in their responses. One example of a response containing the expression is: “*process of elimination, a minus t inverts the signal, and then moved in time*”. Using responses that contained the root of one of the words, they were able to provide information on the uncertainty of students answers. Students who get the multiple choice response correctly, but mention either “guess” or “eliminate” do not demonstrate full understanding of the concept being tested. Figure 3 shows both the prior and post tests, and comparison of correct and correctly guessed responses.

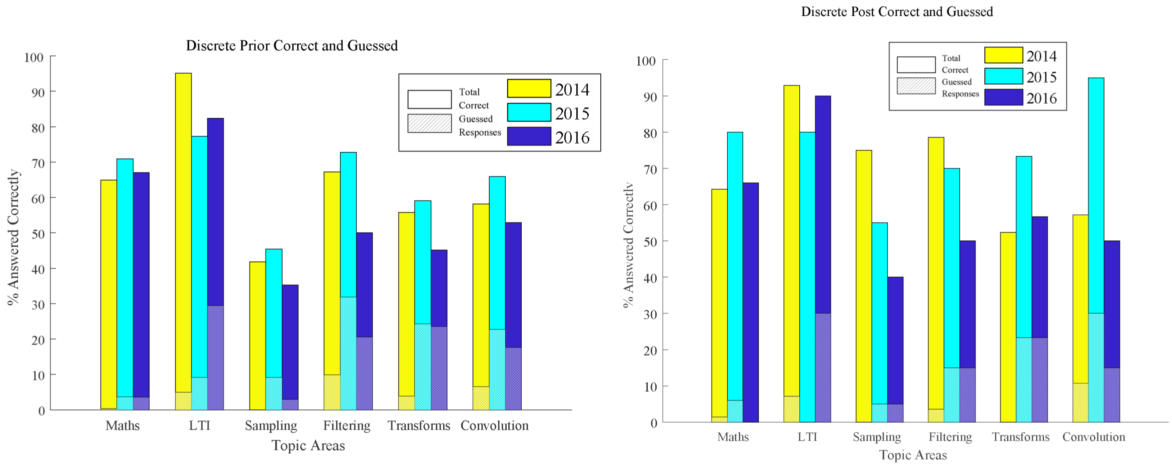

Figure 3: Correct Vs. Guessed Responses

From these figures, we can quickly identify which areas have high levels of correct response, but also high levels of guessed responses. This provides further valuable information for identifying areas where high level of confusion may be present.

For most topics, and most years, there is a decrease in the percentage guessed responses from the prior to post test. This is reassuring, as the concepts tested are reinforced throughout the teaching of the unit. However, there is a few topic areas which the number of guesses is high in the post test. This is most likely due to the small sample sizes for some post test cohorts.

Multiple Choice and Text

Further links between the multiple choice responses and textual responses were investigated. Question 7 in the discrete concept inventory was selected to be investigated closely, due to its high number of incorrect responses. Using the software Leximancer, the correct textual responses for this question were analysed using the parameters discussed above.

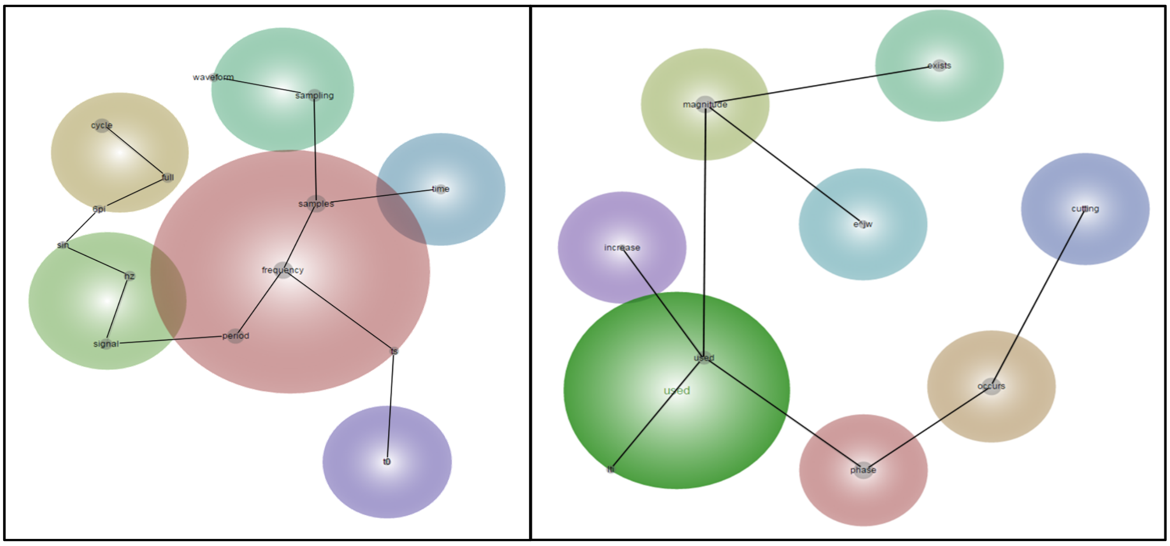

Figure 4(a and b): Q7 (left) compared with Q8 (right) concept maps

Figure 4a shows the links between several key concepts. Each key concept identified can be seen within a circle, and the size of that circle represents its significance, based on frequency counts . The distance between the concepts represent how closely they were mentioned in conjunction with one another. The main focus of these figures however, is their shape.

Compared below in Figure 4b is the correct responses for Q8. Q7 had a high incorrect rate, but a relatively low guess rate. However, Q8 had a much migher correct rate, but in turn a much higher guessing rate. Both Q7 and Q8 fall in the sampling topic area being analysed. Looking at each of these concept maps, it can clearly be seen that the one on the right demonstrates a much more disjointed representation. This is believed to be due to the high number of responses guesssed Q8, meaning there was not one common link between concepts contained within the question, as revealed by students' textual explanations.

Conclusions and Recommendations

Having access to the multiple choice response, as well as textual reasoning gives a much richer and informative insight into students' conceptual understanding. Comparing three years of data informed instructors to place more emphasis in teaching on the low-scoring concept areas (such as the topic area of sampling). The added text component also allows conceptual understanding to be assessed more definitively when compared to the single multiple-choice selection. As a starting point, areas which have high guessing rates have also been identified.

As previously stated in the research questions, several similar and different aspects were able to be identified between the cohorts. Deduced from the data analysed, persistent misconceptions are identified (such as the concept area of sampling). These are identified in two ways, firstly from high levels of guessing, and secondly from low scores derived from the multiple-choice selections. Adding the textual analysis component allows for not only those responses guessed to be identified, but allows for conceptual understanding to be explored. This is evident in concept maps generated for two selected example questions in the discrete-time signal processing concept inventory.

Once persistent misconceptions are identified, specific teaching strategies can be deployed to assist students in gaining the correct conceptual understanding. This links back to the second research question. For example, considering student difficulties in the area of sampling, a number of teaching strategies can be used. One approach relies on presenting more examples that describe the relationships between the continuous signal and its sampled versions, at various sampling rates. These examples can be supported by both mathematical and graphical representations. Further, audio signals with varying bandwidths (for example, speech and music signals) can be sampled at different rates, and the results played to assist students (audibly) identify the practical effects of varying the sampling rates on the outcome. In addition, the effects of varying the sampling rates on the signal representation in the frequency (Fourier) domain can be used to provide another way of demonstrating those effects analytically and visually. Using the examples as starting points for class discussions is another strategy that can help with enhancing conceptual understanding.

Future work will include investigating more automated lexical analysis techniques to extract further concepts and information about the student's understanding from the given text response. This will hopefully lead to further insights for both students as individual learners, and educators for the class as a whole. Using textual analysis can help see on an individual scale to find one students misconceptions, and on a whole group scale to find class based misconceptions.

References

Arbogast, C., & Montfort, D. (2016). Applying Natural Language Processing Techniques to an Assessment of Student Conceptual Understanding. 2016 ASEE Annual Conference & Exposition Proceedings. doi:10.18260/p.26262

Birenbaum, M., & Tatsuoka, K. (1987). Open-ended versus multiple-choice response formats—it does make a difference for diagnostic purposes. *Applied Psychological Measurement*, 11(4), 385-395.

Goncher, A. M., Jayalath, D., & Boles, W. (2016). Insights into students' conceptual understanding using textual analysis: A case study in signal processing. IEEE Transactions on Education.

Libarkin, J, L. (2008). Concept Inventories in Higher Education Science. A manuscript prepared for the National Research Council Promising Practices in Undergraduate STEM education Workshop 2. Washington, D.C., Oct. 13-14, 2008.

Pellegrino, J, Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: Committee on the Foundations of Assessment, National Academy Press.

Popping, R. (2012). Human or Machine Coding of Open-ended Questions. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 115(1), 79–88. doi:10.1177/0759106312445710

Prevost, L.B., K.C. Haudek, E.N. Henry, M.C. Berry, & M. Urban-Lurain, (2013). Automated text analysis facilitates using written formative assessments for Just-in-Time Teaching in large enrollment courses. 2013 American Society for Engineering Education (ASEE) Conference & Exposition, Atlanta, GA.

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. The authors would also like to thank Dr Dhammika Jayalath for his efforts in implementing this concept inventory.