

# Using automated text analysis to evaluate students' conceptual understanding

*Andrea Goncher, Wageeh Boles and Dhammika Jayalath  
Queensland University of Technology*

*andrea.goncher@qut.edu.au, w.boles@qut.edu.au, dhammika.jayalath@qut.edu.au*

---

## **BACKGROUND**

A major challenge for assessing students' conceptual understanding of STEM subjects is the capacity of assessment tools to reliably and robustly evaluate student thinking and reasoning. Multiple-choice tests are typically used to assess student learning and are designed to include distractors that can indicate students' incomplete understanding of a topic or concept based on which distractor the student selects. However, these tests fail to provide the critical information uncovering the how and why of students' reasoning for their multiple-choice selections. Open-ended or structured response questions are one method for capturing higher level thinking, but are often costly in terms of time and attention to properly assess student responses.

## **PURPOSE**

The goal of this study is to evaluate methods for automatically assessing open-ended responses, e.g. students' written explanations and reasoning for multiple-choice selections.

## **DESIGN/METHOD**

We incorporated an open response component for an online signals and systems multiple-choice test to capture written explanations of students' selections. The effectiveness of an automated approach for identifying and assessing student conceptual understanding was evaluated by comparing results of lexical analysis software packages (Leximancer and NVivo) to expert human analysis of student responses. In order to understand and delineate the process for effectively analysing text provided by students, the researchers evaluated strengths and weakness for both the human and automated approaches.

## **RESULTS**

Human and automated analyses revealed both correct and incorrect associations for certain conceptual areas. For some questions, that were not anticipated or included in the distractor selections, showing how multiple-choice questions alone fail to capture the comprehensive picture of student understanding. The comparison of textual analysis methods revealed the capability of automated lexical analysis software to assist in the identification of concepts and their relationships for large textual data sets. We also identified several challenges to using automated analysis as well as the manual and computer-assisted analysis.

## **CONCLUSIONS**

This study highlighted the usefulness incorporating and analysing students' reasoning or explanations in understanding how students think about certain conceptual ideas. The ultimate value of automating the evaluation of written explanations is that it can be applied more frequently and at various stages of instruction to formatively evaluate conceptual understanding and engage students in reflective learning.

## **KEYWORDS**

Learning analytics, automated lexical analysis, conceptual understanding

---

## Introduction

Conceptual understanding of fundamental STEM ideas is critical to developing students' ability to solve problems and apply knowledge in different contexts (Rittle-Johnson et al., 2001; Streveler et al., 2008). Many university-level engineering students still have low conceptual understanding of fundamental engineering concepts at the completion of their courses due to misconceptions limiting or preventing conceptual change (Streveler, Brown, Herman, & Montfort, 2014). While barriers to learning difficult concepts and preventing conceptual change can be difficult to overcome, Perkins (2007, p.45) offers several heuristics for educators including going "beyond the topic to the symptoms" and further addressing the symptoms to get to the causes.

A major challenge for assessing students' conceptual understanding of STEM subjects is the capacity of assessment tools to reliably and robustly evaluate student thinking and reasoning. Multiple-choice tests are often used to assess student learning and understanding, especially for large class sizes, but do not provide measures of higher-level thinking. Another drawback to multiple-choice testing is that it lacks the capacity to assess a fully accurate understanding of concepts. For example, a student may select the correct answer from four or five possibilities, but not have a complete understanding of that conceptual area. In this case, the student may have an accurate understanding to the point of eliminating incorrect responses, but the multiple-choice style of test does not evaluate how the student would explain why that selection is correct compared to a process of elimination or guessing. Deficient mastery of a conceptual area can also allow misconceptions to persist as the student continues to learn and build on prior knowledge. Previous studies on conceptual understanding in STEM subjects used short-answer questioning, i.e. questioning that requires more explanation, to reveal that students held both correct and incorrect ideas, which went undetected by multiple choice testing (Prevost, Knight, Smith, Urban-Lurain, 2013).

Concept Inventories are a type of assessment that aims to evaluate students' conceptual understanding by developing questions as well as the answer selections to reveal the types of misconceptions. Incorrect answer selections, or "distracters", are developed to reveal students' alternative conceptions. 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 (Buck & Wage, 2006). The SSCI was developed to assess five conceptual dimensions and distracters were designed to highlight common misconceptions or rote memorization, i.e. being able to use forward as well as reverse reasoning. While the SSCI was created and validated to identify student misconceptions, it still lacks the capability to capture students' thought processes and reasoning that can potentially uncover why a misconception exists and persists. We hypothesize that the addition of a textual response component, where students can explain or provide reasoning as to why a selection is correct, will provide the data needed to analyse the root causes of misconception rather than identifying only the symptoms of misconception.

The goal of this study was to evaluate lexical analysis methods for assessing open-ended responses. We chose to evaluate three approaches based on the benefits and challenges to efficiently and effectively assess large numbers of textual assessment data. Ultimately, the findings from this research will inform how lecturers can use students' written responses to multiple-choice questions to evaluate students' reasoning without the time consuming process of individually marking student assignments. Analysing and evaluating the capacity and effectiveness of lexical analysis approaches is a first and crucial step to be able to understand and evaluate students' conceptual understanding.

# Text Analysis

Text analysis includes both qualitative and quantitative methodologies and can employ three methods of coding: Manual, computer-assisted, and computer-generated codes (Leximancer, 2011). Text mining, a form of qualitative analysis, involves the extraction of key concepts and the categorization of these concepts into themes, i.e. a family of concepts, nodes. In this paper we utilize and analyse these three methods to evaluate how concepts are extracted from open response textual data. The three methods to extracting and categorizing concepts have different requirements and subsequently various strengths and weaknesses.

## Approach

This applied study used manual, computer-assisted and automated techniques text analysis approaches to identify the concepts or conceptual areas that students used to explain their multiple-choice selections to questions assessing some elements of signals and systems conceptual understanding. We incorporated an open response component (in addition to the multiple-choice component) for a subset of questions from the SSCI to capture students' explanation of their selections. The multiple-choice and open response questions were trialled on undergraduate electrical engineering students (N=60) participating in a digital communications unit/subject. We tested students on 15 questions selected from the 25-question Signals and Systems Concept Inventory. These questions covered five fundamental conceptual areas. The question distribution among those conceptual areas is: Foundational mathematics, Linearity & time invariance, Transform representations, Convolution and Filtering.

The effectiveness of an automated approach for identifying and assessing student conceptual understanding was evaluated by first defining and refining the coding processes and then comparing results of lexical analysis software packages (Leximancer and NVivo) to expert human analysis of student responses. Our research framework, illustrated in Figure 1, outlines the three approaches and the processes for analysing the data set based on textual responses to the selected questions from the SSCI. In order to understand and delineate the process for effectively analysing text provided by students, the researchers evaluated strengths and weakness for both the human and automated approaches.

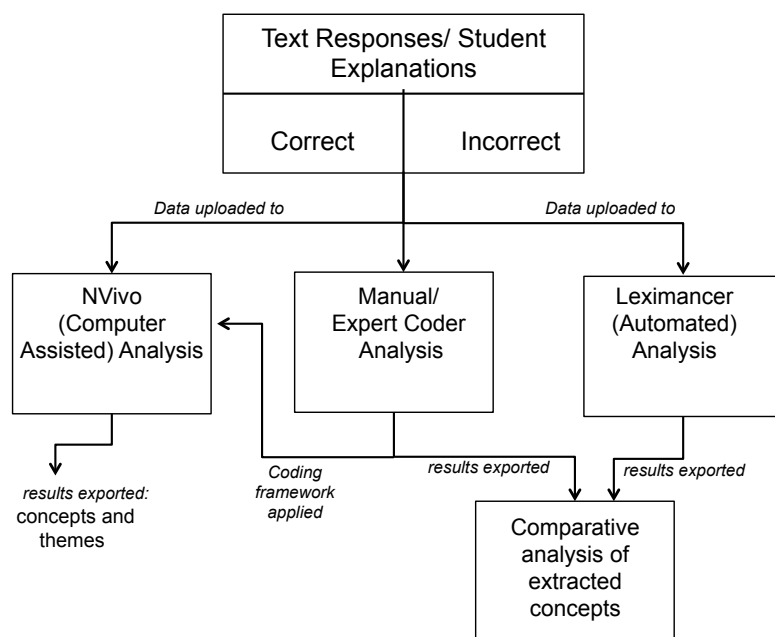


Figure 1. Research framework

# Analysis

## Manual Analysis

An expert coder in the area of theme identification and development was selected to analyse the data set in order to extract concepts, unbiased by the content material. Selecting a manual coder with expertise in the process of identifying concepts rather than the content area is more similar to the computer-assisted or automated approaches chosen in this study. The manual analysis of the student quiz free text responses was predominantly inductive and interpretive, without theoretical modelling. These emphases were based on the objective to identify the key concepts held by the students, as revealed in their free/ open text explanations.

### *Inductive*

Patterns in the data were identified in a “bottom-up” approach, the text of the responses determining the themes identified. The coding schema was not devised before analysis and the concepts discovered were linked directly with students’ own statements. This approach contrasts with deductive processes, which determines themes first, which are identified in the data (Braun & Clarke, 2006).

### *Interpretive*

The concepts identified were based on the apparent meaning of the students’ statements, rather than just the vocabulary used. This required an interpretation of the sense of the text written by the students. This approach contrasts with a semantic level of analysis, which is based purely on the words used by the students. (Braun & Clarke, 2006)

While the meaning of the students’ statements was sought, associated vocabulary for each concept was identified, towards the wider goal of developing a computer-based response system. This revealed a fundamental and problematic difference in approach between human and automated text analysis.

### *Process*

The text responses were treated question-by-question and separating those associated with correct and incorrect multiple-choice answers (See “Manual/ Expert Coder Analysis,” Figure 1). A five step iterative process was followed, until no significant further insights were found. This process was repeated, until stable themes had been identified.

1. Overview of responses: All responses relevant to a specific question and correct/incorrect response group were read through, as a familiarisation stage.
2. Note recurring concepts and phrases: Concepts, which were common across the text, were noted, including key phrases which typified the responses.
3. Review for less obvious themes: Statements, which did not readily fit into the recurring concepts, were reviewed to consider adding further core themes.
4. Group statements: Statements were grouped into key concepts areas.
5. Read through, testing themes: The statement groups were tested against the whole data set relevant to that question and correct/incorrect grouping.

## Leximancer

The automated lexical analysis software, *Leximancer*, had parameters and settings to account for stopwords, e.g. “the”, “and”, “is”, synonyms, and proper nouns. However, the settings required manual input and refinement based on the imported data set.

The textual responses associated with incorrect and correct answers were separated into two excel files and marked by the SSCI question as well as the participant ID. Our overall

unit of analysis was structured by question (rather than participant) for this stage of the study in order to identify the types of concepts present in the student responses for a conceptual area. Focusing on the question, or conceptual area initially will provide a set of themes that we can use to apply to future algorithms that can identify potential misconceptions in new textual data, e.g. new student responses.

We analysed the set of data collected from students' written explanations (textual responses) (See "Leximancer (Automated) Analysis", Figure 1):

1. The whole data set, using a standard set up, as installed
2. The whole data set, using a revised set up, refined through trial and error
3. A data set separating correct and incorrect responses, using the revised set up determined in 2).

## Findings

### Manual Analysis

The core concepts identified are presented in Table 1, with representative quotes and the vocabulary observed as being associated with each concept. Questions 1-4 in Table 1 illustrate the types of concepts that were present for each respective question. We also provide the concept and distractors intended to test the students that were developed as part of the SSCI. The core concepts identified for the sample questions were not developed in a specific order, though the more obvious/better supported ones are listed first.

Table 1. Identified concepts from manual coding

	<b>Question 1</b>	<b>Question 2</b>	<b>Question 10</b>	<b>Question 13</b>
<b>Concept</b>	Sinusoidal frequency	Signal delay; plot shift	Transform Representations	Convolution
<b>Distractors</b>	High amplitude/ large period	Shift direction/ distance	Faster oscillation/ lower amplitude; involves less than 2 cosines	Square pulses in the time domain; Frequency/ amplitude change for filter
<b>Concepts extracted from manual coding</b>				
<i>Core concept 1</i>	Time: period	Shift: left/right, plot, back/forward, negative/positive	Differences: two frequencies (lower frequency has greater magnitude), main/carrier/fundamental signal has low frequency and other/noise/harmonic signal has higher frequency; not symmetrical	Sameness: no difference, identical transform, perfectly match, no change
<i>Core concept 2</i>	Size: peaks/troughs, amplitude	Time: delay, lag	Sameness: same frequency, same magnitude, same amplitude; mirror images, line up	Plot: triangle, ramp up and down

<i>Core concept 3</i>		Result: delay, decrease	Plot: shape, look	Magnitude: amplitude; twice, double, added, omega doubled; not changed
<i>Core concept 4</i>	Formula: $2\pi w_0, f=1/T$	Formula: $p[n-2]$	Varying: middle (has more magnitude)	Frequency: multiplication
<i>Other responses</i>	Guess	Only one aspect changes	Guess	Guess; Period: change, same

## Leximancer

We used the results/ output from the manual and automated analysis (See “Comparative analysis”, Figure 1) to evaluate the effectiveness and accuracy of an automated approach when compared with a purely manual approach. The standard/ base-line settings in Leximancer analysis on the total data set did not result in useful themes. Adjusted results on the (a) whole data set and (b) separated correct/incorrect subsets did not result in a better match for manually identified concepts and automated concept extraction, as detailed below.

- a) Whole data set (correct and incorrect data combined, i.e. whole data set, standard set-up; whole data set, revised set-up.
  - i. Machine coding was similar to manual coding in at least one theme – 73% of questions.
  - ii. Machine coding was dissimilar to manual coding in at least one theme – 100% of questions.

The first statistic offers some indication that machine coding may provide correlation with manual coding, however the second statistic reveals that machine coding always disagrees to some extent with manual coding. From this, one may conclude that when a correlation exists, it is not a strong one.

- b) Separated correct/incorrect responses
  - i. Correct/incorrect analyses overlapped – 80%
  - ii. High ranking themes and associated terms were based on meaningless vocabulary – 30% of questions
  - iii. Overall, themes and associated terms were based on meaningless vocabulary – 87% of questions

The first statistic indicates that machine coding provides limited means of distinguishing correct from incorrect conceptualisations. For example, the machine analysis for Question 1 returns “frequency” with 100% connectivity as an explanation for both correct and incorrect quiz responses (connectivity is an indication of the connectedness of concepts within the theme and is a measure of the importance of the theme in the data set); similarly for Question 7 “period” is returned with 90% connectivity as an explanation for both correct and incorrect responses. The second and third statistics indicate that machine coding is not based on meaningful terms in the text (despite attempts to configure the software to help it do this consistently). For example, for Question 13 “best” is given as a theme with 100% connectivity; similarly, for Question 10 “unfortunately” is given as a theme with 100% connectivity (ahead of a relevant and valid term such as “amplitude” with 73% connectivity).

There were several single word themes, which did not have any associated vocabulary:

- a) At least one theme produced for the question – 100%

b) Themes produced which had perfect connectivity – 67%

The first statistic indicates that machine coding identifies themes based on limited meaning scope. The second statistic indicates that the machine considered a significant proportion of these limited scope themes to be highly defensible. The settings were configured to try to alleviate this, however it is probably a result of the limited amount of data provided.

The ability of the software to provide consistent results across hardware contexts was unreliable based on our analysis. An alternative strategy for using Leximancer included entering and perhaps prioritising correct concepts. This was more useful in guiding the software to form themes, however it also reduced the likelihood of revealing incorrect concepts held by the respondents. This is potentially a problem for overlooking the types of concepts held by a smaller population of students.

## NVivo

NVivo provides a powerful and flexible management tool for qualitative analysis and reporting. Some automated coding is offered, however this is dependent on manual configuration of either source documentation formatting before import into NVivo or data querying within NVivo. The pattern-based auto coding in NVivo 10 is still in the experimental phase (QSR International, 2014) and was designed to facilitate the coding process for large volumes of textual data. Auto-code can be applied to existing patterns, where the software will compare a sentence or paragraph to the content already coded at the previously defined codes. We found this problematic for our data set because responses were often short (1-2 sentences) for each entry, so it was challenging to develop patterns to apply to new entries. This is also a problem for uncovering new types of misconceptions held by students that were not previously identified in the stable coding structure.

For this analysis, the data was imported and coded according to a) student respondent, b) association with correct/incorrect quiz answers and c) categorisation of tested concepts. The manual coding of the data question-by-question and by correct/incorrect quiz answers is possible, however it would require entire manual coding.

To examine the current data set and the available types of concepts we applied the framework developed from manual coding to the data set structure the types of concepts present in the set of student responses. The following codes in Table 2 show the developed framework from the manual coding and structure of the SSCI. The coding could be combined and compared usefully using queries offered in NVivo; however the foundational coding and set up, as well as the querying process, was entirely manual. In summary, subject to further investigation, NVivo did not appear to be a useful tool for carrying out a completely automated textual analysis.

Table 2 NVivo Codes

<b>Parent Nodes</b>	<i>Concepts</i> <i>Correctness</i> <i>Participant</i> <i>Question</i>			
<b>Child Nodes</b>	Nodes: <i>Concepts</i>	Nodes: <i>Correctness</i>	Nodes: <i>Participants</i>	Nodes: <i>Question</i>
	<ul style="list-style-type: none"> <li>• Convolution</li> <li>• Filtering</li> <li>• Linearity &amp; Time Invariance</li> <li>• Mathematics</li> <li>• Transform</li> </ul>	<ul style="list-style-type: none"> <li>• Correct</li> <li>• Incorrect</li> </ul>	<ul style="list-style-type: none"> <li>• Participant ID number</li> </ul>	<ul style="list-style-type: none"> <li>• Q1-Q15</li> </ul>

	Representations and Fourier			
--	-----------------------------	--	--	--

## Benefits and Challenges of Each Approach

Based on the textual data, use of manual, computer assisted, and automated approaches applied in this study, we present Table 3 below of the main identified benefits and challenges to identifying the types of concepts present in students' explanations.

Table 3 Type of approach and its applicability

<i>Approach</i>	<i>Challenges</i>	<i>Benefits</i>
<b>Manual</b>	Time consuming to apply to large volumes of textual data sets.	Ability to accurately identify new types of concepts and potential outliers not identified by the coding structure.
<b>Computer-Assisted</b>	Auto-codes are based on predefined codes and depend on matching patterns (can't identify new types of codes automatically)	Minimize time/effort of manual input and can be applied to a well-defined coding structure.
<b>Automated</b>	Parameters required manual setup for each new imported data set.	Apply concept extraction strategies to large volumes of textual data sets.

## Conclusions

This study evaluated the usefulness of incorporating computer assisted or automated coding to analyse students' reasoning or explanations in understanding how students think about certain conceptual ideas. We identified the benefits and challenges of using the various approaches to analysing textual data and found that the automated input of some approaches was both beneficial and problematic. A predefined coding structure, developed by experts, is accurate and can be refined but is more time consuming and less flexible when implemented into a computer-assisted approach. The investigations we conducted so far show that the three types of approaches all have benefits and challenges to using them to identify the types of concepts present in students' explanations or reasoning to answering multiple-choice questions.

Determining the benefits and challenges of various approaches to text analysis empirically using student data is an important step to developing an efficient and effective method for analysing students' understanding of difficult concepts. In our next phase of research, we plan to integrate the useful outcomes of each approach and try to address the challenges by refining our analysis.

The ultimate value of this research on automating the evaluation of students' written explanations is that it can be applied more frequently and at various stages of instruction to formatively evaluate conceptual understanding and engage students in reflective learning. This approach to formative assessment aims at reducing the resources required to evaluate higher-level thinking and increasing the capacity to ensure enhanced student learning. With validated tools and approaches to accurately identify and assess student understanding, we can help lecturers to identify and address misconceptions formatively for large class sizes.

## References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.



- Perkins, D. (2007). Theories of difficulty. In Student learning and university teaching, eds. N. Entwistle and P. Tomlinson. British Journal of Educational Psychology. Monograph Series II: Psychological Aspects of Education.— Current Trends.
- Prevost, L. B., Knight, J., Smith, M. K., & Urban-Lurain, M.. (2013). Student writing reveals their heterogeneous thinking about the origin of genetic variation in populations. In National Association on Research in Science Teaching. presented at the 04/2013, Rio Grande, Puerto Rico: NARST.
- QSR International. (2014). The Nvivo Blog. Retrieved from:  
<http://blog.qsrinternational.com/experiment-with-pattern-based-auto-coding/>
- Rittle-Johnson, B., Seigler, R.S. & Alibali, M.W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93(2). 346-362.
- Streveler, R.A., Litzinger, T. A., Miller, R.L., & Steif, P. S. (2008). Learning conceptual knowledge in engineering: Overview and future research directions. *Journal of Engineering Education*, 97(3), 279-294.
- Streveler, R., Brown, S., Herman, G. L., & Montfort, D. (2014). Chapter 4: Conceptual change and misconceptions in engineering education: curricular-, measurement-, and theoretically-focused approaches, *Cambridge Handbook of Engineering Education Research*. Eds. A. Johri and B. Olds. Cambridge University Press.
- Wage, K. E., Buck, J. R., and Hjalmarson, M. A. (2006). "Analyzing Misconceptions Using the Signals and Systems Concept Inventory and Student Interviews," in Proc. 4th Signal Processing Education Workshop, September 2006, pp. 123–128.

## Acknowledgements

This study was sponsored as part of Learning and Teaching grant through Queensland University of Technology. Special thanks are due to all the student participants for their feedback and Ian Stoodley for his working developing and analyzing the coding structure.

## Copyright statement

Copyright © 2014 Goncher, Boles and Jayalath: The authors assign to AAEE and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to AAEE to publish this document in full on the World Wide Web (prime sites and mirrors), on Memory Sticks, and in printed form within the AAEE 2014 conference proceedings. Any other usage is prohibited without the express permission of the authors.