



Quantext Quick Reference Guide
Pre-release 2017

Access Quantext

Logging into Quantext	
Url	http://quantext.co.nz
Login with Twitter or Google	Once you are on the Quantext homepage you can login with either Twitter or Google. You must already have a Twitter or Google (e.g. gmail) account.
Support	To contact support or to report any errors. Please email: info@quantext.co.nz

Getting your data into Quantext

Manage student data files	
Upload a new student file	Choose the browse button to locate the file you want to upload (.xlsx) then click <i>Upload</i> . Refer Quantext Data File Format Guide for details of the required file formats, limitations and caveats.
Current student files	Click to view the files you have uploaded, ordered by newest file first. Select one or more files for analysis.
Past analyses	You can analyse more than one data file at a time. This is useful for doing comparisons. Give the current analysis a name and you can run it again with the same files without having to reselect them.
Manage reference corpora	
Upload new reference corpus	Choose the browse button to locate the teaching file you want to upload (.pdf) then click <i>Upload</i> . Refer Quantext Data File Format Guide for details of the required file formats, limitations and caveats.
Current reference corpora	Click to view the names of files you have uploaded. Ordered by newest file first.
Current analysis	
Name	Enter a name for the analysis. After you run the analysis, the files selected for the analysis will be saved so you can run again using Past analyses .
Current student files selected	The files selected from listed current student files.
Current reference corpus	The file selected from listed reference corpora.
Run this analysis	Run the analysis with the files selected.

Analysing data with Quantext

General settings	
Change question	Select the question to analyse from the file you uploaded.
Edit blacklist	Add any word you like to the default blacklist. Remove any word you like from the default blacklist. See also: Exclude blacklist words... IN PROGRESS
Customise dictionary	NOT YET IMPLEMENTED
Customise punctuation list	The default punctuation list at present is: . , ; () \ / _ @ # ^ NOT YET IMPLEMENTED
Current reference answer	The currently selected reference answer for the question. By default there are no reference answers. You can add as many reference answers as you like using the Edit answers dialogue box. The current reference answer is then selected from the list you have created.
Edit answers	Add a new reference answer to the list. The ability to edit or delete answers from the list is NOT YET IMPLEMENTED
Edit categories	Create labels or categories to assign to student responses. For example, a student response that corresponds to a model answer might be assigned the label, <i>correct</i> . In order to assign the <i>correct</i> label to any student response you first need to create the label using Edit categories . Categories or labels are listed, as they are created and preceded by an ⊗. Clicking on the ⊗ will delete that label and it will no longer be available to select in the <i>Category</i> column in the spreadsheet view.
Top keywords & phrases – select to filter in spreadsheet view	
Number of keywords/phrases to display	Select from 10 to 30 keywords and phrases to display. This is currently available in steps of 5. i.e. 10, 15, 20 etc.
Display bigram or trigram phrases	Bigrams consist of two words. Trigrams consist of three words. Collectively, they are called ngrams. You can select whether to display either key bigrams or key trigrams. i.e. these are groups of two or three words that occur together more often than expected by chance.

Exclude blacklist words	Exclude blacklist words from student responses. Blacklist words are typically function words like a, the, my, your etc. The default Quantext blacklist contains more than 1000 common function words. Use Edit blacklist to add or remove words from the default blacklist.
Filter punctuation	Remove punctuation listed in the punctuation list .
Normalise contractions	Convert contractions like, <i>don't</i> , <i>can't</i> to their normal form. i.e. <i>do not</i> , <i>cannot</i> . This can be useful when you want to combine word or ngram counts. For example, some students might write <i>don't</i> , others might write <i>do not</i> . You are probably interested in the combined frequency of these expressions rather than having them separated. In this case, it makes sense to select normalise contractions .
Correct spelling	Apply spelling correction to student responses. NOT YET IMPLEMENTED
Filter numbers	Filter numbers from student responses.
Transform to lower case	Convert all responses to lowercase. This is recommended to count for example <i>The</i> and <i>the</i> in the frequency count of <i>the</i> .
Ngram measure	Different statistical measures can be used to calculate the most frequent ngrams (bigrams and trigrams in Quantext). Four different measures are provided in addition to a simple <i>raw</i> frequency count: Likelihood Ratio (LR), Pointwise Mutual Information (PMI), Students t-distribution (STUDT), Chi square distribution (CHISQ).
CG cutoff	Relates to a collocation graph of responses. NOT YET IMPLEMENTED
Window	Relates to a collocation graph of responses. NOT YET IMPLEMENTED

Readability Indices – Select to show	
Response length (words)	Total number of words per response.
Sentences/response	Number of sentences per response.
Lexical diversity (TTR)	Lexical diversity provides a number between 0-1 which gives the type to token ratio (TTR) - the number of unique words divided by the total number of words. For example, ' <i>The cat sat on the mat</i> ' has a TTR of $5/6 = 0.83$
Lexical density (LD)	A number between 0-1 which indicates the ratio of content words to the total number of words. Content words are nouns, verbs, adjectives and adverbs.
SMOG	One of many readability indices which roughly correlate with the level of educational attainment, SMOG is included in Quantext as an example. It is commonly used to assess the readability of important health information materials. SMOG stands for 'Simple Measure of Gobbledygook' and the range of values is typically between 5 - 18. It roughly correlates with reading level expected after 5 - 18 years of formal education. Like many indices, it is only valid where responses are a reasonable length, i.e. at least 30 words.
Similarity	<p>Similarity of student response to the selected reference answer or 0.0 if no reference answer is selected. Similarity ranges from -1 (completely opposite meaning) through 0 (not similar) to 1 (same meaning).</p> <p>Similarity to the reference answer is calculated using a pre-trained word2vec model of word embeddings. We hope to add further models to Quantext over time and provide the ability to customize models.</p>

Reference corpus	
Reference corpus	Click on this tab to display summary data for the reference (teaching) corpus file (if uploaded). Clicking on any top listed keywords or keyphrases from the student responses will produce a <i>Keyword in context</i> (KWIC) display of that keyword/keyphrase where it occurs in the reference corpus.