Tool for the Automatic Analysis of Lexical Sophistication
(TAALES)

User Manual for TAALES 1.4 (updated 11-15-2015)
Kristopher Kyle and Scott Crossley
Georgia State University

This document is intended to assist users of TAALES. It includes a brief explanation of how to use the tool. Additional information about TAALES is included in the supplementary Index Description Spreadsheet (available at www.kristopherkyle.com).

Please use the following citation when referencing TAALES in your work:

Kyle, K. & Crossley, S.A. (in press). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*. doi: 10.1002/tesq.194

Options

TAALES analyzes a number of textual aspects related to lexical sophistication. The user can choose to include indices related to all of these aspects, or can choose to constrain the output.

The user may choose to include indices from any (or all) of the following categories:
- word frequency and range
- n-gram frequency
- psycholinguistic information
- academic language

Input

All input files must be text files (.txt) that do not include any type of markup (e.g., XML, HTML, etc.). Files must be located in a single folder. TAALES will process all .txt files in the chosen input folder.

Saving Your Output

TAALES provides output in the form of a comma-separated (.csv) file that can be opened with any spreadsheet software. The default output file name is "results.csv", though we would recommend changing this file name each time you run TAALES to ensure that the file is not overwritten.

Indices

TAALES 1.4 calculates 104 indices in five categories: Word Frequency, Range, N-grams, Academic Lists, and Word Information. All indices in TAALES 1.4 are normed. Please see the supplementary spreadsheet file for more information. Also, see the article referenced above for more information.