Rules for Spec Templates script  
  
Goals:

* Use samples and test templates to develop and test functionality.
* Process header and footer and document margins and styles to match from input file to output file (keep existing processing for now – we will refine this later)
* Take input word document (SECTION 26 05 00.docx or SECTION 26 05 29.docx) and update formatting to match the test template’s (test\_template.docx) list formatting.

Assumptions:

* Input doc will have all content formatted as list items except:
  + Header and footer (keep existing processing for now)
  + Section number – always one line of text “SECTION” followed by the section# sometimes formatted XXXXXX sometimes XX XX XX
  + Section title – one line of text
  + End of section – one line of text “END OF SECTION” sometimes followed by the section# sometimes formatted XXXXXX sometimes XX XX XX
  + Sometimes blank lines or single lines of text at the end of the document for Page intentionally left blank
  + Some tables may not be formatted in list items – these should just be included in the previous list item and formatting should be kept
  + Some incorrectly formatted content might have manually numbered sections. These should be included in the output doc with special care taken to attempt to reinsert them back into the correct list levels. They will most likely have the correct numbering manually entered as content, like “C. This is content for line number C”. If the list level cannot be determined they should be included in the output file but marked by highlighting and included in an error report.
* All items in the input doc that are not the SECTION TITLE AND NUMBER can be any string of text and should not be assumed to be specific text. The correct list level cannot be determined by searching for specific text.
* All list levels in the test\_template.docx will have a label or name that begins with “BWA”. These should be recreated to be used in the output document.

Strategy:

* Analyze the input document to determine the list level first using list items detected by word.
* Pair list level and list level numbering with blocks of content
* Determine if the list levels follow the correct hierarchy inherent to all lists.
* Check if numbering follows the correct format from the template
* Attempt to correct broken lists by searching content blocks for numbering and use that to reconstruct the list
* Output content blocks that cannot be processed or formatted with styling like highlighting to help debug process