Numeracy App data Script

The automated data script for the Numeracy app was made using Python 3.9 and will likely not work on older version. I would suggest opening the script in an editor such as PyCharm with a python 3.9 environment to run it (at the time of writing this, psychopy is still using Python 3.6 which will not work).

The script has two parts:
Data_Extraction_Numeracy.py – This is the main script that loops through the files
Numeracy_App_Script.py – This is the script that extracts the data from a single file

You only need to run the *Data_Extraction_Numeracy.py* file which will call a function in the *Numeracy_App_Script.py* for each file in the folder. Once it's done it will then save all the data into an excel sheet called "Numeracy_Data.xlsx" in a "Results" folder. If these last two do not exist it will create a results folder (in the same directory as the script) and the excel file.

The folder should look something like below, with the scrips and the data in one folder. The data in the data-folder need to be text files, but the naming doesn't matter that much.

…Folder/
- Data_Extraction_Numeracy.py
- Numeracy_App_Script.py
- Data/
  - PP01.txt
  - 02.txt
  - NumeracyPP03.txt

Data_Extraction_Numeracy.py

For the script to work it needs to be in the same folder as the "data" folder, not in the data folder but next to the data folder, because it will try to find a "data" folder in the same directory as itself.

It will proceed by loading in all the .txt files in the data folder and loop through each file, processing it using *Numeracy_App_Script.py*. If the IDE has a debug-window or something similar, a message will appear
`Processing: PP01.txt`

If something goes wrong, the message will indicate which file caused an error
`Error Occurred at file: PP01.txt`

Upon completion, all the data is saved to an excel file in the Results folder. This can be loaded into an SPSS file or processed further in excel.

Sheet	Description
Measures	A key for the description codes
Numeracy Data	▪ A univariate data structure. Each participant has 11 rows, one for each measure. ▪ Grade, Birthday, Sex, and Age are available if filled in, otherwise empty. ▪ Components is the measure, followed by the number of correct answers, the total number of questions in that measure, and the percentage correct.

Numeracy_App_Script.py

This script extract the data for individual files, the main script will call the "numeracy_app_data_runner" function that outputs the data from a single text file.

The first thing it does is check which line contains "Notes" and "Score" which, at the time of writing, denotes where the participant data ends and where the task data ends.

```
Test Name:
Child ID: 03
Session ID:
Grade:
Date Of Birth:
Sex: Female
Age: 5
Played At: 2021-02-11 16:57:05 +0000
Notes:                                    ← find this line
.
.
.
82. 76_WordProb4 Answer: 0 Response: 6
83. 77_WordProb5 Answer: 0 Response: 4
84. 78_Equat6 Answer: 1 Response:
85. 79_WordProb6 Answer: 0 Response: 6

Score                                     ← find this line
Early Numerical Concepts and Language: 11/12
Counting a Subset: 5/6
```

Participant Information is the first block, everything above "notes". It reads this in, transposes the rows into columns, and saves it for the output.

Task data is the second block, between "Notes" and "Score". I made it so the script calculates the scores rather than trusting the output. I'm sure the app creators did a good job, but still. The script sums up the 1 or 0 for each question based on their Measure Component, creating 11 scores.

These are outputted as a dataframe to the main script, which adds it to a larger output.

PP	Grade	Birthday	Sex	Age	Component	Correct	Total	Percentage
02	4		Female	2	CountSub	4	6	0,666667
02	4		Female	2	Equat	5	6	0,833333
02	4		Female	2	MatchNum	10	12	0,833333
02	4		Female	2	NumCncpt	12	12	1
02	4		Female	2	NumComp	6	6	1
02	4		Female	2	NumLine	5	6	0,833333
02	4		Female	2	OrdPos	4	5	0,8
02	4		Female	2	Pattern	6	6	1
02	4		Female	2	SpMeas	14	14	1
02	4		Female	2	Subit	6	6	1
02	4		Female	2	WordProb	6	6	1