

Google Sheets v Microsoft Excel: A Comparison of the Behaviour and Performance of Spreadsheet Users

Karl Mernagh

*School of Business and Humanities
Dundalk Institute of Technology
karl.mernagh@dkit.ie*

Dr. Kevin Mc Daid

*School of Informatics
Dundalk Institute of Technology
kevin.mcdaid@dkit.ie*

Keywords: POP-I.C. end user applications, POP-I.C. web, POP-III.B. spreadsheets, POP-III.B. new language, POP-IV.B user interfaces

Abstract: Summary of Findings

Spreadsheet technology has traditionally been limited to a single user operating in a desktop environment and working in an isolated environment. With the advent of Cloud Computing, a paradigm shift has occurred in the way users utilise the collaborative sharing and communication of their work in both an educational and business environment. The opportunities for people to cooperate on multiple spreadsheets at the same time and in real time have grown significantly. However, the behaviour and performance of users working in this new paradigm has not been explored to a great extent in scientific. In comparison to desktop spreadsheet technologies such as Excel, Cloud based spreadsheet technologies have only started to be investigated in relation to user behaviour and performance.

The purpose of this paper is to examine user interaction with Cloud based spreadsheets compared to desktop based spreadsheets. Specifically, we have focused on the leading technologies in these fields, namely Microsoft Excel and Google Sheets.

This paper investigates the effectiveness and efficiency of user interaction with traditional Excel spreadsheets and Google Sheets using a group of third level business students that have been trained in both applications. The students were randomly assigned to a technology and then asked to complete a list of standard spreadsheet tasks. In both cases, user interaction with the spreadsheets was tracked and recorded on a cell by cell basis using automated technology.

The findings of this paper are that users completed a set of basic tasks assigned to them in a quicker and less error prone manner when working with Cloud based Google Sheets compared to desktop based Excel spreadsheets.

The Importance of Cloud based Spreadsheets

In a paper by Brian Sommer, 2014, "So long, spreadsheets -- hello in-memory cloud financial tools", he identified the following advantages of cloud computing:

Controls over data exist for only a portion of the reporting and data capture systems. Spreadsheet systems are often absent effective dating, restart/recovery technology.

Over time, Financial/Accounting groups have compounded the problem by increasing the number of interrelated spreadsheets. As stated by Brian Sommers, "*They've added baling twine, bubble gum and bandages to the imperfect solution and have created a monster.*"

The problem has gotten even more troubling as new kinds of data are being utilized by businesses and merging these data types into legacy solutions and patchwork coding will only increase the problem.

Cloud based data is particularly safe as the data is stored across multiple servers, but in the off chance that this data is lost there is no way to retrieve it unless a local backup copy is maintained by the end user.

The issue of security and confidentiality is a factor when storing information in the Cloud. With hackers now focusing more of these large datasets it could be very easy to have a data breach that causes large quantities of data to be stolen or corrupted.

Outline of Experiment

This initial experiment was designed to examine the efficiency and effectiveness of spreadsheet usage for the tasks of text entry, data entry and formula entry for both users of Microsoft Excel and Google spreadsheets in a similar format to Olsin & Nilsen, 1987. The objectives of the experiment are specified as follows.

a) Objectives: Investigate efficiency and effectiveness of spreadsheet use for

i. Text Entry.

The data examines the time to enter text and the correctness of the entered data.

ii. Data Entry.

The data examines the time to enter numeric data and the correctness of the entered data.

iii. Formula Implementation.

The data examines the time to implement formulas and the correctness of the formulas such as SUM, COUNT, AVERAGE, MAX, MIN and COUNT

iv. Time Taken To complete Experimental Task

In addition to the elements above, the experiment examined the time taken to complete the entire experimental task from start to finish. The data examines how long it took the participants to complete the experiment from start to finish.

b) Experimental Structure

The experiment consisted of two groups of third level final year business studies students that were given a set of procedures to carry out. The groups were split randomly into those that would complete the set of tasks using Excel and those that would complete the tasks using Google spreadsheets. Both sets of tasks were identical. The students were both provided with a sample file with data that they had to insert into certain cells and use formulas to calculate formulas. The students were given 5 minutes to acclimatise themselves to the exam format and to log into Google spreadsheets using their Gmail accounts while the other group copied an Excel file from a Shared Directory. Once the files were opened the recording of the exam start time began in both cases.

Recording Technology

The recording technology for the Excel spreadsheets used VBA code to track which cells the user selected, which cells they entered data, text and formulas into and when they altered these values. The results were stored in a hidden spreadsheet. It recorded the duration the user stayed within a cell and what they did while in that cell. This technology has been used by a number of other researchers including Bishop (2003) and McKeever (2011).

The recording technology for the Google spreadsheets was limited to a small extent due to the fact that Google spreadsheets will only trigger an event when the user exits a cell. This means that we could track what the user has done inside the cell and when the user left the cell. Google spreadsheets are incapable of tracking when a user enters a cell or how long they stayed within the cell. If the user returns to the cell and makes changes we could record this information with the exception that if the result of their alteration resulted in the same value being produced. This would not trigger a recording action. All of this is because Google wants to reduce the load on the server which is hosting the spreadsheet. So only the OnEdit exit event gets fired for cell changes.

b) Results

A highly noticeable behaviour that was observed with the Excel user group was that a large portion of the group entered their data in a horizontal manner. This means that they filled out the first row of data in its entirety before moving to the next row of data to be inserted. This could explain the difference in performance in text, numerical and formula entry. This did not occur with the Google group. Overall the majority of users entered their data in a vertical manner. This is discussed further in the discussion section.

i. Text Entry

The measured times in seconds to complete the text entry task for the participants using Microsoft Excel and Google spreadsheet technologies is presented in the box plot below. In total 13 participants completed the task using Microsoft Excel and 14 completed the task using Google Spreadsheets. In completing the task all but two of those using Microsoft Excel completed this element of the task completely with all but one doing so using Google Spreadsheets.

While the boxplot shows the completion times for Excel participants to be higher than those for Google, the boxplot shows a single extreme value of 322 seconds in the completion times for Google Spreadsheets. Removal of this value yields a distribution of completion times for Google participants that follow a normal distribution.

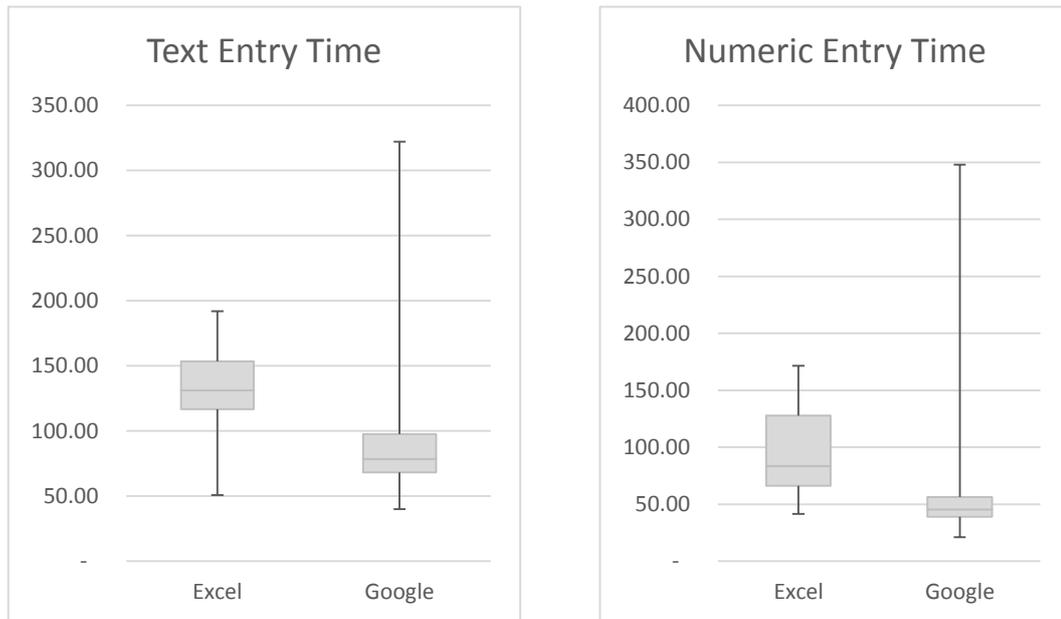
In formally comparing the performance of the two groups using an independent sample approach the one tailed p-value is 0.07. On removal of the extreme Google value the p-value drops to 0.0004, thus providing significant evidence at the 5% level that participants using Google were quicker to complete the task than those using Excel.

ii. Numeric Entry

The measured times in seconds to complete the numeric entry task for the participants using Microsoft Excel and Google spreadsheet technologies is presented in the box plot below. As before, the total number of participants is the same. In completing the task all but two of those using Microsoft Excel completed this element of the task completely with all but one doing so using Google Spreadsheets.

While the boxplot shows the completion times for Excel participants to be higher than those for Google, the boxplot shows a single extreme value of 348 seconds in the completion times for Google Spreadsheets. Removal of this value yields a distribution of completion times for Google participants that follow a normal distribution.

In formally comparing the performance of the two groups using an independent sample approach the one tailed p-value is 0.15. On removal of the extreme Google value the p-value drops to 0.0001, thus providing significant evidence at the 5% level that participants using Google were quicker to complete the task than those using Excel.



iii. Formula Implementation

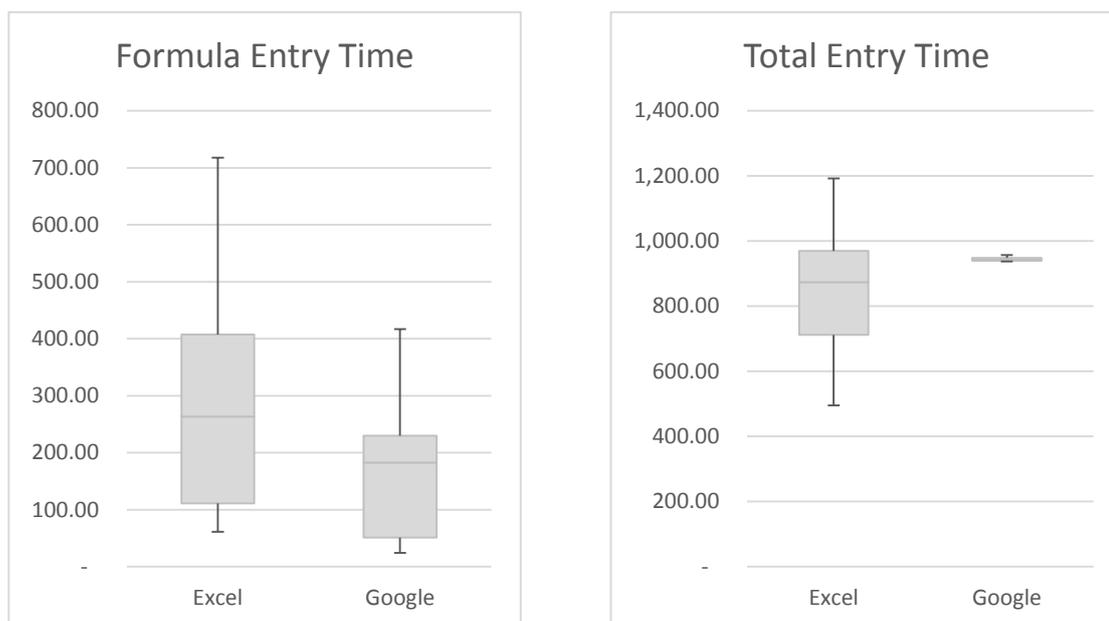
The measured times in seconds to complete the formula entry task for the participants using Microsoft Excel and Google spreadsheet technologies is presented in the box plot below. As before, the total number of participants is the same. In completing the task all but two of those using Microsoft Excel completed this element of the task completely with all but one doing so using Google. The boxplot shows the completion times for Excel participants to be lower than those for Google,

In formally comparing the performance of the two groups using an independent sample approach the one tailed p-value is 0.03.

iv. Time Taken To Completion

The measured times in seconds to complete the experiment for the participants using Microsoft Excel and Google spreadsheet technologies is presented in the box plot below. In total 13 participants completed the task using Microsoft Excel and 14 completed the task using Google Spreadsheets. In completing the task all but two of those using Microsoft Excel completed this element of the task completely with all but one doing so using Google Spreadsheets. Note that the average score out of 20 for Excel users was 19.8 with the average score for Google users of 19.9.

In formally comparing the performance of the two groups using an independent sample approach the one tailed p-value is 0.38.



Discussion

Less user interaction was recorded for Google spreadsheets due to the ability of Google spreadsheets to record every interaction without the use of third part libraries such as the GData library. This was expected due to the limitations currently surrounding recording of interactions with Cloud technology. Cloud technology only records user interaction after a cell has been exited and there isn't a native way to get around this without a third party library. Excel used VBA code to record user interactions which allowed for a much more detailed set of results to be recorded.

The experiments provided some evidence to support the conclusion that participants using Google sheets outperformed those using Excel in so far as they took significantly less time to complete the tasks. Also, in completing the tasks they scored the same on the basic tasks of text and numeric entry but significantly higher for the task of formula implementation. These results are somewhat surprising for two reasons. Firstly, the participants were probably more familiar with Excel than Google technology. Secondly, distributed spreadsheets such as Google sheets are browser based and with this there is a slight delay in the execution of tasks such as entry of data and formulae to cells which might have been expected to adversely affect the performance of the Google sheets group.

There are a number of issues that need to be discussed relating to the generalizability of these results.

Firstly, the participants were students and while their background was in business their knowledge of spreadsheets is most likely not at the same level of real daily spreadsheets users.

Secondly, the task selected was short and, while it included the key elements of text, numeric and formula entry, a longer more involved task may better reflect the real development tasks performed by spreadsheet users.

Thirdly, the Excel user group behaviour of entering data in a horizontal manner had a significant time to completion factor for each set of tasks. The Google user group filled out the spreadsheet in a vertical manner. They worked from column to column, inserting textual data, numerical data and using the fill down functionality for formulaic entry.

We can only speculate that the Excel users had difficulty with absolute cell references as it appears in the recorded data that they made multiple attempts to fill down once they had a formula correct but had not used the appropriate absolute referencing to get the correct answer. The Google groups' data shows that they had no difficulty with cell referencing and that very few errors were made when filling down formulaic entries.

Bearing these points in mind, and considering that this is the first experiment that seeks to compare the performance of users of these technologies, the authors feel that this research provides some support for the argument that using Google Sheets is as least as effective as using Excel to develop spreadsheets but that further work is needed to provide stronger evidence.

Conclusion

In conclusion, this initial experiment of comparing user behaviour and performance using a Cloud based spreadsheet technology versus traditional Excel technology has produced a surprising set of results.

Users that are more familiar with Excel seem to have behaved in a less efficient manner. Users that had limited experience and exposure to Google spreadsheets outperformed their Excel rivals in all tests recorded. As this is a small sample group using only two technologies, it is worthy of further investigation in both a larger user base and a more experienced user base.

Also worth noting is that the spreadsheet error checking technologies are in the main not compatible with Cloud based spreadsheets which leaves both market gap and more significantly for industry an absence of error checking. This is important as Microsoft Excel 365 has moved to a Cloud based environment.

As applications such as Office 365, which are pushing users into the Cloud, emerge and are adopted more widely by industry it would be interesting to see if Excel in the Cloud can match Google Spreadsheets or one of the other Cloud based spreadsheet applications that are emerging at present.

Another point that would need to be studied further is the programmability of these applications in a Cloud based environment. Without the use of the GDATA API, Google spreadsheet programming abilities are severely hampered. This issue may become less relevant as Google expands the capabilities of its programming environment with new iterations of their technology. Office 365 has a vast amount of programmability using technologies ranging from the REST API to HTML5 and XML to implement VBA but will they suffer the same deficiencies as Google Spreadsheet programming.

References

- [1] Sommer, B. (2014). So long, spreadsheets -- hello in-memory cloud financial tools
<http://www.zdnet.com/so-long-spreadsheets-and-hello-in-memory-cloud-financial-tools-7000026176/>
- [2] Leonard, J. (2013) Office 365 and Google Apps fight for supremacy
<http://www.computing.co.uk/ctg/analysis/2269524/office-365-and-google-apps-fight-for-supremacy>
- [3] Rust, A., Mc Daid, K., Bishop, B. (2008) Investigating the Potential of Test-Driven Development for Spreadsheet Engineering
<http://arxiv.org/ftp/arxiv/papers/0801/0801.4802.pdf>
- [4] McKeever, R., (2011) How do Range Names Hinder Novice Spreadsheet Debugging Performance?
<http://arxiv.org/ftp/arxiv/papers/1009/1009.2765.pdf>
- [5] McKeever, R., (2011) Effect of Range Naming Conventions on Reliability and Development Time for Simple Spreadsheet Formulas
<http://arxiv.org/ftp/arxiv/papers/1111/1111.6872.pdf>
- [6] Olsin, JR & Nilsen, E., (1987) Analysis of the Cognition involved in Spreadsheet Software Interaction, Human-Computer Interaction, Volume 3 Issue 4, December 1987, Pages 309-349