

What Is the Best Way For Developers to Learn New Software Tools?

A Small Empirical Comparison Between a Text and a Video Tutorial

Verena Käfer¹, Daniel Kulesz¹, and Stefan Wagner¹

¹Software Engineering Group, Institute of Software Technology, University of Stuttgart, Germany

Abstract

The better developers can learn software tools, the faster they can start actually using them and the more efficiently they can later work with them. Tutorials are supposed to help here. While in the early days of computing, mostly text tutorials were available, nowadays software developers can choose among a huge number of tutorials for almost any popular software tool. However, only little research was conducted to understand how text tutorials differ from other tutorials, which tutorial types are preferred and, especially, which tutorial types yield the best learning experience in terms of efficiency and effectiveness, especially for programmers.

To evaluate these questions, we converted a “proven” video tutorial for a novel software tool into a content-equivalent text tutorial. We then conducted an experiment in three groups where 42 undergraduate students from a software engineering course were commissioned to operate the software tool after using a tutorial: the first group was provided only with the video tutorial, the second group only with the text tutorial and the third group with both.

Surprisingly, the differences in terms of efficiency are almost negligible: we could observe that participants using only the text tutorial completed the tutorial faster than the participants with the video tutorial. However, the participants using only the video tutorial applied the learned content faster, achieving roughly the same bottom line performance. We also found that if both tutorial types are offered, participants prefer video tutorials for learning new content but text tutorials for looking up “missed” information.

We mainly gathered our data through questionnaires and screen recordings and analyzed it using t-tests or Wilcoxon signed-rank tests for finding significant differences and a Shapiro-Wilk test to see if a group follows a normal distribution. The data is available at [14].

Since producing tutorials requires effort, knowing with which type of tutorial learnability can be increased to which extent has an immense practical relevance. We conclude that while it would be ideal if software tool makers would offer both tutorial types, it is more efficient to produce only text tutorials – provided you manage to motivate your learners to use them.

1 Introduction

Typical developers have to work with many different tools every day. While many developers get frustrated when the tools’ user interfaces do not match their expectations [18], most developers manage to come to terms with what they get. However, as technology evolves, developers are expected to get along with new tools quickly. Thus, the learnability and understandability of a software tool is an important success factor [7, 8]. Some software tool producers believe to tackle these issues appropriately by providing tutorials for the software. However, as Martin et al. have shown, tutorials are often unavailable, incomplete or focusing on the wrong aspects [20].

While in the early days of computing mostly text tutorials were available [29], advances in technology made it possible to easily produce video tutorials and even interactive tutorials. Van Loggem’s research indicates that the “classic” written (and printed) manuals usually are not the first choice – at least for most tool end-users today [30]. Instead, most tool users prefer interviewing colleagues or searching for

tutorials in online sources, and especially the latter offer many different kinds of tutorials to choose from. Furthermore, many tutorials are not crafted by the original makers of the software or hired tutorial producers but by other tool users who post such tutorials on blogs or social media platforms.

In this study, we focused on passive tutorials, meaning tutorials which present information but do not react to the user in any way, as they are the most common ones for application software.

1.1 Problem Statement

A literature review showed that there are some papers which address the differences between paper-based tutorials and video tutorials. Interestingly, there are many different results (see section 2.2). What we wanted to know is which tutorial types are preferred by **developers** and especially which tutorial types yield the best learning experience for them in terms of efficiency and effectiveness. This is a problem because developers are overburdened with too many choices while software tool makers do not know on which type of tutorials they should spend their (typically) limited resources. All these issues can be combined into one research objective.

RO: What is the best way for developers to learn new software tools?

1.2 Research Objective

To address our research objective, we formulated the following research questions:

- RQ1: What kind of tutorial do developers prefer if both text and video tutorials are available?
- RQ2: Which tutorial takes learners less time?
- RQ3: Which tutorial is more effective?

1.3 Context

The study was conducted using 42 students from an undergraduate software engineering course. We used an experimental software for testing spreadsheets which is implemented as an add-in for Microsoft Excel [22]. Since both the tool and its underlying approach are novel, learning and understanding them requires adequate tutorials. Therefore, results obtained from this study should be considered especially in contexts where developers are confronted with new software tools which are not directly mappable to previous experience.

The study described in this paper borrows its design and task descriptions from another study conducted in early 2016 by the second author (the study has not been published yet). However, it is not a pure replication: while the original study only used video tutorials to explain SIF, this study also used content-equivalent text tutorials which were produced for this very study. Furthermore, the participants were asked more questions about their perceptions on the tutorials they consumed, while this aspect was not targeted in-depth in the original study.

This paper follows the structure proposed by Jedlitschka et al. [10] with a slight deviation (we describe the procedure of the experiment earlier than proposed by Jedlitschka et al. and merged the alternative technologies and the related studies): after describing the background of the study, we explain the plan of the experiment and its deviations, we present the obtained results and discuss them before drawing a final conclusion and outlining future work. The graphical representations are based on the recommendations by Tufte [26].

2 Background

2.1 Technology under Investigation

The *Spreadsheet Inspection Framework* (SIF) [12] is a software application for faults in spreadsheets which has been developed at the University of Stuttgart and already has been evaluated in a number of previous studies [13, 16, 15]. It is operated through a plug-in for Microsoft Excel which allows end-users to automatically scan spreadsheets for “bad smells” (e.g. formulas referencing the same cell twice) or violations of design rules (e.g. formulas with constants). Apart from running these pre-defined scans, it also allows end-users to specify their own test scenarios which are comparable to unit tests for normal software applications. While SIF is primarily targeted at end-users, we expected developers to understand

the tool's concepts faster, because these concepts are similar to static analysis and unit tests in traditional software development.

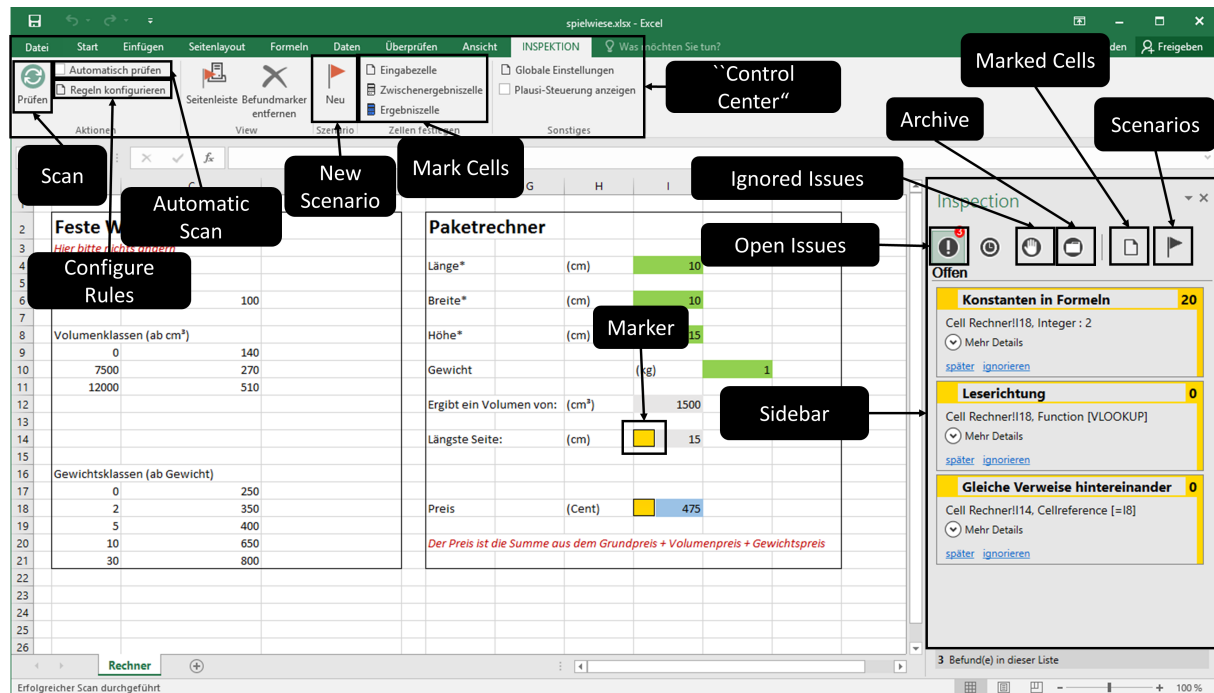


Figure 1: Main user interface of the Spreadsheet Inspection Framework

A screenshot of the main user interface of SIF is provided in Fig. 1. It shows that SIF adds a new tab on the ribbon bar which acts as SIF's "control center". Here, the user can configure the rules to be used for testing spreadsheets, create new test cases and start the automated scans of the spreadsheet. The sidebar on the right side provides the user with an overview about open, postponed, ignored and solved findings. Furthermore, the issues reported in the sidebar are synchronized with marker symbols in the corresponding cells of the spreadsheet. Although none of these elements of the user interface is overly complex, learning how to use them all in just a few minutes can be challenging for end-users – making this setting a proper ground for evaluating the efficiency and effectiveness of different tutorials.

For the purpose of another study [17], the second author has produced an introduction video and two video tutorials (educational screencasts) which explain the ideas behind SIF and how to use it. Since we regard comparing tutorials with different (depth of) content as not being fair, the first author simply converted the video tutorials into content-equivalent text tutorials, making them directly comparable. This also ensured that the videos had the same origin and thus were comparable to "official tutorials" produced by an original software maker (unlike tutorials made by hired tutorial producers or third parties).

2.2 Related Studies

Overall, there are only a few contradictory studies which investigated tutorials for software. On the one hand, a study by Mestre [21] indicated that learners favored a static text tutorial over video tutorials. On the other hand, several studies showed that learners prefer video tutorials over text tutorials [5, 28, 19]. On top of that, a study by Palmiter [23] came to the conclusion that both text tutorials and video tutorials have their specific benefits. And there are even studies that could not find significant differences between text and video tutorials at all (e. g. [24, 3, 6]).

Our impression from these studies is that both video and text tutorials have their particular advantages and disadvantages. For video tutorials Alexander reported that learners watch the whole video(s) (which is time consuming) [3] instead of picking particular scenes. However, studies by Mestre and Alexander both confirm that learners have trouble finding specific scenes after watching a video [21, 3]. As an advantage of video tutorials, participants in a study by deVaney perceived them as "more human" [6]. For text tutorials, the study by Mestre came to conclusion that two huge advantages are that (i) the participants can decide which part of it they want to read and (ii) that information in a text can be found

faster than in a video [21].

Overall, it was not possible to find a common denominator regarding to which tutorial type is the best one. Additionally, none of the mentioned studies investigated programmers in particular. The aim of our study was to fill this gap.

2.3 Practical Relevance

The better users can learn software tools, the faster they can start actually using it and the more efficiently they can later work with it. Thus, good learnability is a vital success factor for software tools. Therefore, knowing with which type of tutorial learnability can be increased has an immense practical relevance.

Furthermore, producing tutorials also requires effort. In our own experience, we perceived producing video tutorials to be far more difficult than producing text tutorials (the second author worked about 80 hours to produce three video tutorials with a total length of less than 25 minutes). Therefore, knowing if video tutorials actually have advantages in terms of learnability is also relevant for practice.

3 Experiment Planning

3.1 Goals

For every research question we defined a goal:

- Goal 1: Analyze video and text tutorials
For the purpose of understanding which tutorial type is preferred
With respect to the preference rate of the users for each type
- Goal 2: Analyze video and text tutorials
For the purpose of comparing their efficiency
With respect to the time required by users to complete the tutorials and the tutorials plus a following tasks
- Goal 3: Analyze the quality of video and text tutorials
For the purpose of comparing their effectiveness
With respect to counting how often users looked something up, the number of wrong actions they took, measuring the perceived difficulty level of the tutorials and with respect to which percentage of the learning content they understood right away

Depending on the context, the meanings of the terms efficiency and effectiveness can vary. Therefore, we want to give our definition of these terms in the context of this paper. An explanation how we measured them will be presented in section 3.7.

Efficiency Efficiency is the ratio between input and output. In our case, we investigated the ratio of the content of the tutorials to the time needed to consume this content. The faster a tutorial teaches some content, the more efficient it is.

Effectiveness Effectiveness measures to what degree a goal was reached with a given method. In our case, there are two types of effectiveness: (I) effectiveness in terms of using a tutorial as a source to look up information and (II) effectiveness in terms of using a tutorial to learn how to solve a task. Therefore, the easier information can be looked up in a tutorial and the better it teaches how to solve a task, the more effective it is.

3.2 Participants

To successfully pass one of the undergraduate software engineering lectures, all students were required to take part in one study, having the choice between three other studies and ours¹. We decided to use this population primarily because it was easily accessible. There were 42 participants and they were about 20 years old. Nine of them were female and thirty-three were male.

¹The other studies were about coupled file changes, statical code analysis and feature development.

We gave the participants the choice between 14 time slots for doing the experiment, reserving them on a “first-come-first-serve” basis. Once three participants reserved a time slot, the time slot was full because we ran three parallel experiments in each time slot (nevertheless, the experiments were strictly separated as described in section 3.3).

Each participant had to fill out an online questionnaire a few days before the experiment. Based on the completion order of this questionnaire we assigned each participant a unique anonymous ID.

To ensure confidentiality, each participant had to sign an agreement at the beginning of the experiment about not disclosing any details of the experiment to other participants. Additionally, we stressed this point at the beginning of each experiment and explained the motivation behind this. All participants were allowed to abort the experiment at any time if they wished but no participant used this opportunity.

3.3 Procedure

Figure 2 provides an overview of the procedure in our study. The actual steps where participants use tutorials are highlighted in light gray. However, we simplified the “final test” in this figure to make it more concise – in reality, some parts of the questionnaire were dependent on the type of experiment the participant had.

As it can be seen, we randomly divided the participants into six groups. Every group contained seven

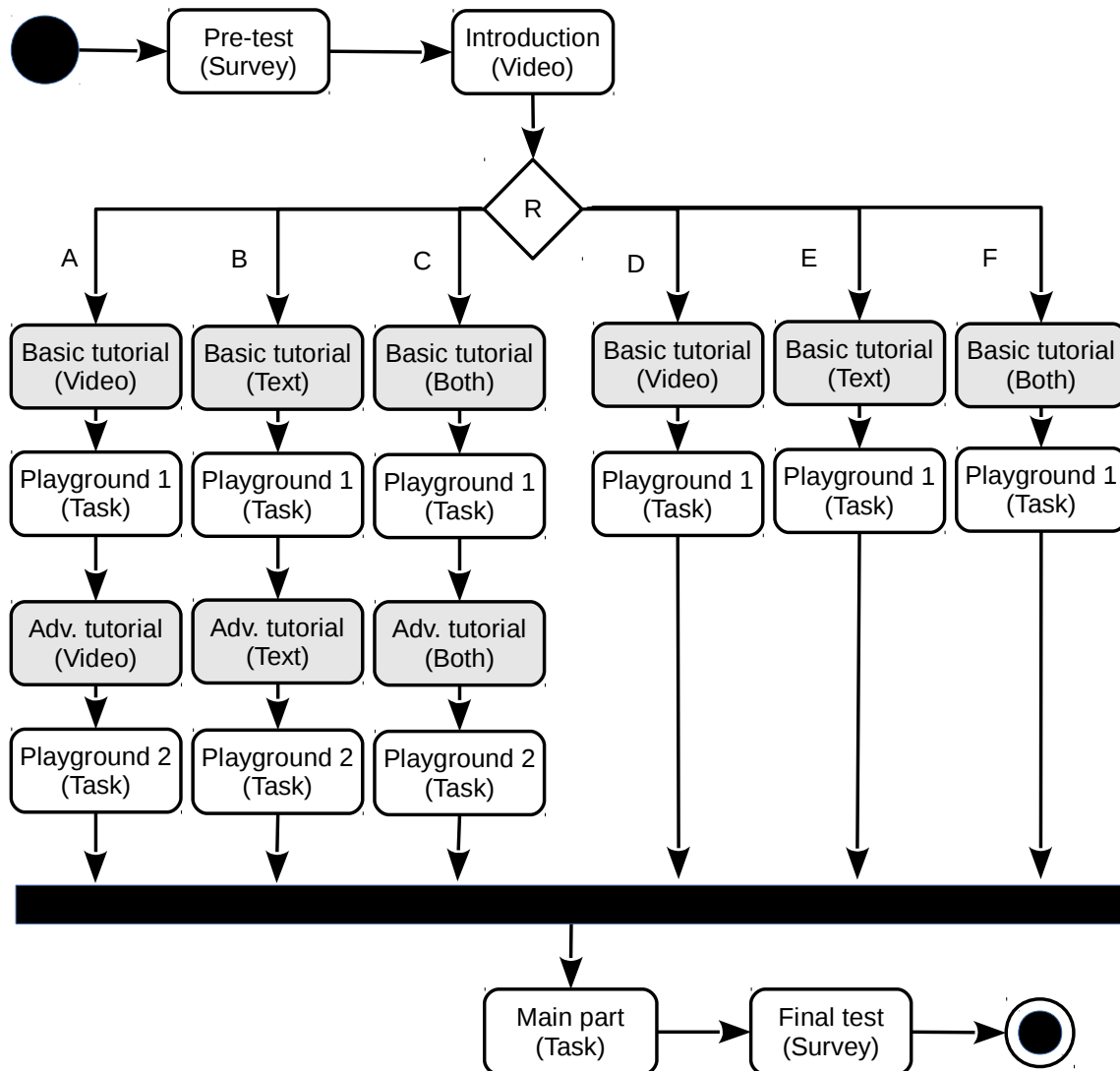


Figure 2: Overview of the experiment procedure (slightly simplified)

participants.

- A: with scenarios, video only
- B: with scenarios, text only
- C: with scenarios, video and text
- D: no scenarios, video only
- E: no scenarios, text only
- E: no scenarios, video and text

All participants had to watch the introduction video. After using their particular tutorial, all participants solved the first task (Playground 1). After finishing the first task, the participants of groups A, B, and C directly continued with the main task. By contrast, participants of the other groups first had to do another round of tutorials in which they learned the (advanced) scenario testing technique and solved simple tasks with it (Playground 2) before continuing with the main task.

Since each of our time slots had the capacity for three participants, we either executed experiments A, B and C or D, E and F (the idea behind this was not to give participants the feeling that they are slower than others if they see that others who started at the same time had already finished). The particular assignment to A, B or C or D, E or F respectively was based on the order of arrival of the participants.

Before the actual start of the experiments, we explained to each group of three participants the whole setup and what they had roughly to do, not giving them any details except “you will get instructions where everything will be explained”. The experiments took place in one of our computer pools with air-conditioning, on standard desktop computers with 23” monitors with a full HD resolution. The “work-places” of the participants were close to each other but we put separating walls in-between as shown in figure 3, so the participants could only see their individual screen.



Figure 3: Our computer pool where the experiments took place

For later examination, we recorded all experiments using a screencast recorder (due to legal privacy restrictions we recorded only the computer screen and not the participants). Also, we watched the participants' screens over VNC connections during the experiment thus avoiding to directly look over their shoulders so as to not make them feel uncomfortable. This was a tough trade-off, as it also made it impossible for us to count how many times the participants looked something up in the text tutorials ourselves.

There were some common issues which most participants encountered, such as not knowing how the VLOOKUP formula works in Microsoft Excel. In such cases we gave the participants "meta hints" like "maybe you could try to look this up on the internet?". We allowed access to the internet because it is a very common way to retrieve information about unknown formulas. We also had a look at what they looked up on the internet and can confirm that it was nothing irrelevant to our study.

SIF also has several known bugs and we were aware of them already before the experiments (they are not straightforward to fix). If the participants encountered such a bug we helped them get around it. We took extreme care not to disturb the other participants in such cases as we did not want to influence them.

3.4 Experimental Material

The participants of our experiment had to learn how to use the *Spreadsheet Inspection Framework* (SIF). The introduction video had a length of 09:03 minutes. The video about the basics of SIF was 05:42 minutes long and the scenario video had a length of 10:01 minutes. The introduction video explained general problems of spreadsheets and the concept of design patterns for spreadsheets, whereas the other two videos explained the tool SIF itself (what does it do), how to use static rules and how to create a scenario. The videos were not chunked or split in chapters. They did not cover all functions of SIF but all functions that were necessary to solve the given tasks.

Following existing studies [25, 21, 19, 27], we decided to have videos with a screen recording and narration. In the videos we use two different techniques similar to the ones used by Ilioudi et al. [9]. Either the left side of the screen showed a screen-cast of a spreadsheet being manipulated and the right side showed the second author explaining what he does in the spreadsheet (Talking Head Style) or the whole screen showed the screen-cast of a large spreadsheet with an explaining voice in the background (Khan Style). In every case the clicked buttons and important parts of the GUI were highlighted with blue borders and arrows. An example can be seen in Fig. 12 in appendix A.1. The participants were in full control of the video, meaning they could use every option given by the Windows Media Player.

The text tutorial counterpart for this excerpt of the video is illustrated in Fig. 13 in appendix A.1. It follows the same explanatory scheme as the video tutorial. As it can be seen, similar highlighting techniques have been used in the text to explain which actions the user has to take to reach a certain functionality.

In the pre-test, we asked several questions using an online questionnaire to investigate the participants' experience and opinions. The pre-test was made up of a questionnaire which contained questions about prior knowledge of Microsoft Excel and what tasks the users typically solve using it, if any. It can be seen in appendix A.2.

During the experiment, each participant received printed instructions on what to do next (obviously, the instructions differed depending on which experiment the participant had). The instructions also contained questions about the last task they completed, so the participants worked their way through the instructions, alternating between reading the instructions, doing activities on the computer and writing down answers on the instruction sheets. An example of the instructions is shown in Fig. 4.

Apart from the instructions, we also provided the participants with the (printed) text tutorials, the tutorial videos or both (depending on the group). After finishing all practical tasks, each participant was asked to fill out a paper questionnaire. Here, we asked the participants to self-estimate what percentage of their tutorial they understood when they consumed it for the first time. We also asked them as how difficult they perceived the whole study. Finally, we asked more questions about their background, what they liked and disliked about SIF and, of course, about the tutorials. An example of the final questionnaire for group C can be seen in appendix A.3.

Last but not least the participants were provided with either two or three spreadsheets (depending on the group) which contained a number of seeded errors in order to solve the actual tasks.

Task 4.1

1. You find the latest version of Petra's Excel file in the folder „Study“ on the desktop („rates.xlsx“). Open the file and look around a little bit.
2. Play around in the file, for example change the calculation to „Manager Smart“ (cell B11 in the dashboard)
3. Select the rules “Constants in Formulas”, “Reading Direction” and “Repeated References”
4. Start the analysis. How many issues are there?

Answer:

Rule „Constants in Formulas“: ____ Issue(s)

Rule „Reading Direction“: ____ Issue(s)

Rule „Repeated References“: ____ Issue(s)

Rules for Scenarios: ____ Issue(s)

5. Try to solve all shown issues, but if you couldn't do it after 5 minutes, stop. Could you solve the issues?

Answer:

- Yes, everything worked fine
- Yes, but I think I took longer than 5 minutes
- No, but I could have done it in 20 minutes
- No, I could not do it even in 20 minutes

Figure 4: Translated instructions for the final task

3.5 Tasks

Depending on the group of the participants they had to perform either two or three bigger tasks during the experiment. These bigger tasks were split up into a series of small sub-tasks.

The first bigger task (Playground 1) was to apply the basic functions of the SIF. First, the participants were asked to open a spreadsheet, to activate several static rules and to initiate a scan. For this we prepared a spreadsheet that could be used to calculate the price of a parcel based on its size and weight. If done right, SIF reported for this spreadsheet an issue with the reading direction which could be solved by moving the content in one of the cells. The second issue reported by SIF was that one formula referred to the same cell multiple times using the MAX function (The formula was: MAX(I4;I6;I8;I8;I8;I8;I8;I8;I8)). This issue could be solved by removing the obsolete references (it makes no sense to refer to the same cell multiple times in a formula which is composed only of the MAX function).

The second task (Playground 2) was only performed by participants in the groups that had received the advanced training where they were taught how to use the SIF scenario testing technique. The participants were first asked to create a new scenario on their own and then to use it to find errors in a new spreadsheet. Again we used the example of the parcel price. The created scenario reported an issue because for the final price calculation the weight of the parcel was subtracted. The participants solved this by repairing the formula.

The last and lengthiest task was to extend a given spreadsheet by adding new data and features. The given spreadsheet calculated the monthly bill for using a mobile phone based on a user's consumption of

minutes and text messages (with different rates based on the destination) for different tariffs (which had different rates for these minutes and text messages). It also featured a dashboard where monthly bills could be compared between the tariffs to find the cheapest one. The participants first had to add a new rate and then add the costs for on-net texts for every rate.

3.6 Hypotheses

We wanted to test the following hypotheses. The numbers refer to the research questions.

- H_1 : The video tutorial is chosen more or less often than the text tutorial when both are provided.
- H_{2a} : The three tutorial groups differ in the time they need to complete the tutorials.
- H_{2b} : The three tutorial groups differ in the time needed to complete the tutorials and the following tasks.
- H_{3a} : The participants look up more or less items in the video tutorial than in the text tutorial.
- H_{3b} : The participants differ in their understanding of the video and the text tutorial at the first attempt.
- H_{3c} : The three tutorial groups differ in the number of wrong answers given during the tasks.
- H_{3d} : There is a difference in how difficult the participants thought the tutorial was.

3.7 Analysis Procedure

To analyze the results, we examined the screen recordings and the questionnaires. In the screen recordings, we measured when a participant started a task (when he or she opened the Excel file) and when the participant finished the task (closing the file). The time between one file and the next one was the time used for the tutorial. We measured the duration in seconds and combined the groups with the same used tutorial for the SIF part.

For evaluating the correctness of the results, we designed unit tests with input and expected output values – one unit test for each playground and two unit tests for the final task. We had two unit tests for the final task because the final task included two tasks – the the new rate and then the on-net texts. We then filled in the input values and checked if the actual output values matched the expected values. We judged a unit test for a spreadsheet to be correct if and only if *all* actual output values matched their expected values.

For the questionnaires, we first used LimeSurvey [1] to produce an electronic version of the questionnaires. Then we exported them to R [2] where we evaluated them.

We first determined with a Shapiro-Wilk test if the samples were normally distributed. As the sample groups were quite small, we then used either a t-test (normally distributed) or a Wilcoxon signed-rank test (otherwise) to see if the group results were significantly different. For this, we chose a significance level of 0.05. Additionally, we measured the effect size with Cohen's d and the absolute mean difference MD.

To measure the normal distribution in section 5.3 we used a Shapiro-Wilk test due to a small sample size with a significance level of 0.05.

To measure efficiency, we analyzed the time the participants needed to consume the tutorials. Independent from that, we analyzed the time the participants needed to complete the corresponding playground tasks. We decided to take these separate measures because we wanted to investigate the time distribution between consuming the tutorials and solving the tasks.

To measure effectiveness, we had to consider both types of effectiveness. For the first effectiveness type (looking up information) we counted how often the participants looked something up in the tutorials to see which tutorial is more useful for this task. We would rate the tutorial which was used more often as more effective. For the second effectiveness type (learning how to solve a task) we analyzed the participants' estimates regarding their understanding of the tutorials and their perceived difficulty of the study. Apart from these subjective measures, we also counted the number of wrong actions they took – the less wrong actions they took, the better.

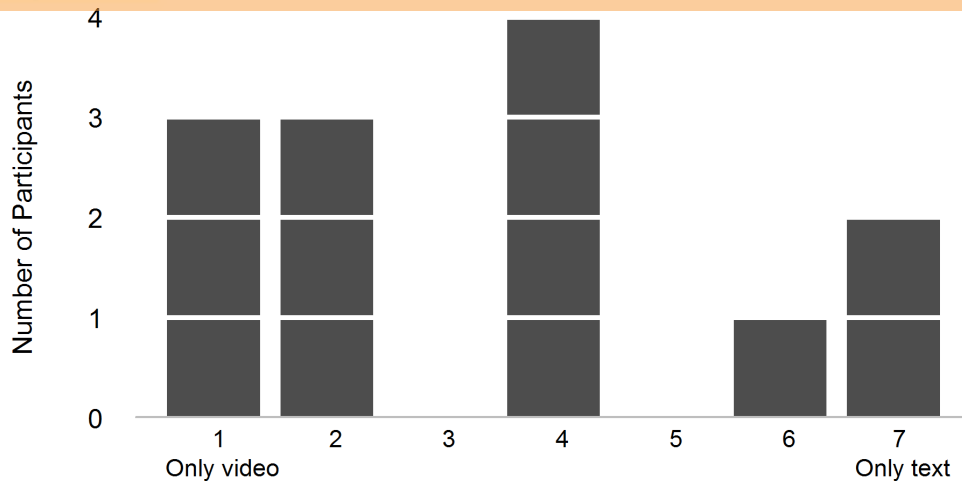


Figure 5: Question for the participants with both tutorials: “which tutorial did you use more often?”

4 Execution

4.1 Preparation

Before the participants arrived, we prepared the computers and started the screen broadcast over the VNC connection. For the participants, there was no particular preparation as all required information was provided during the experiment via the instruction sheets and the particular tutorial. Therefore, we did not ask the participants to prepare themselves in advance.

4.2 Deviations

There were only two deviations from the plan:

- By mistake, we used one of the computer accounts twice, so that the data of the previous participant was overwritten. However, thanks to the screen recording we were able to redo every single action taken by this user in the experiment, recreating the lost spreadsheet.
- One of the participants had issues with his e-mail account so he did not get our invitation for his particular time slot. Therefore, we ran only the remaining two experiments in this time slot. Then, we simply added a new time slot where this participant did the missed experiment alone.

5 Analysis

5.1 Descriptive Statistics

5.1.1 Goal 1 - Usage of the Tutorials

For this goal we asked the groups with both tutorials which tutorial they used more often. Fig. 5 shows the results. As can be seen, from the 13 participants in group C and F who answered this question, six participants used (nearly) only the video tutorial, four used both equally and only three used (nearly) only the text tutorial.

5.1.2 Goal 2 - Duration of Tutorial Usage

Fig. 6a shows that for the SIF tutorial there were only small differences between the time needed by the video and the text tutorial. The participants with both tutorials took more time to complete the tutorial. The standard deviation of all three groups is close around the mean with the deviation of the *video and text* group slightly larger. ($\sigma_{Video}=117.361$, $\sigma_{Text}=139.764$, $\sigma_{Both}=188.432$).

Furthermore, it can be seen in Fig. 6b that there were only small differences in the overall time needed for the tutorial and the following task. The values spread wider around the mean ($\sigma_{Video}=362.353$, $\sigma_{Text}=450.317$, $\sigma_{Both}=344.713$).

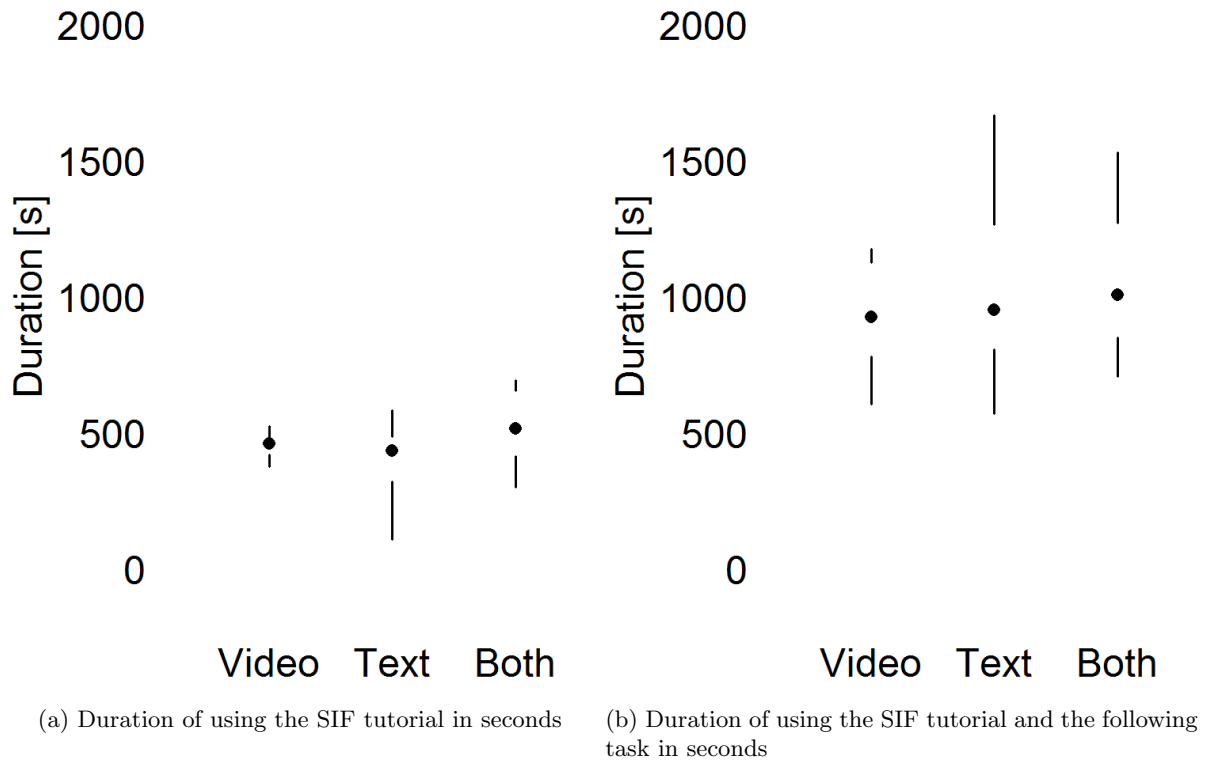


Figure 6: Duration of using the SIF tutorial and using the SIF tutorial plus the following task in seconds

For the second tutorial the results were slightly different. Fig. 7a shows that this time the text tutorial was much shorter than the other two tutorial types. The values are much closer around the mean than for the SIF tutorial ($\sigma_{Video}=362.353$, $\sigma_{Text}= 103.714$, $\sigma_{Both}=54.705$). But again there are close to no differences when comparing the tutorial and the following tasks as can be seen in Fig. 7b. Again, the values are further away from the mean ($\sigma_{Video}=396.333$, $\sigma_{Text}= 357.612$, $\sigma_{Both}=126.839$).

5.1.3 Goal 3 - Effectiveness of the Tutorials

Fig. 8a shows how many times something was looked up in a tutorial by the two groups with text or video tutorial. The figure shows that the participants seldom looked something up in the video tutorial whereas they looked up things many times in the text tutorial.

Fig. 8b shows how many times the participants looked something up if they had both tutorials. Some participants used only one of the two tutorials (the other one was *not used*). Again, they looked up more things in the text tutorial.

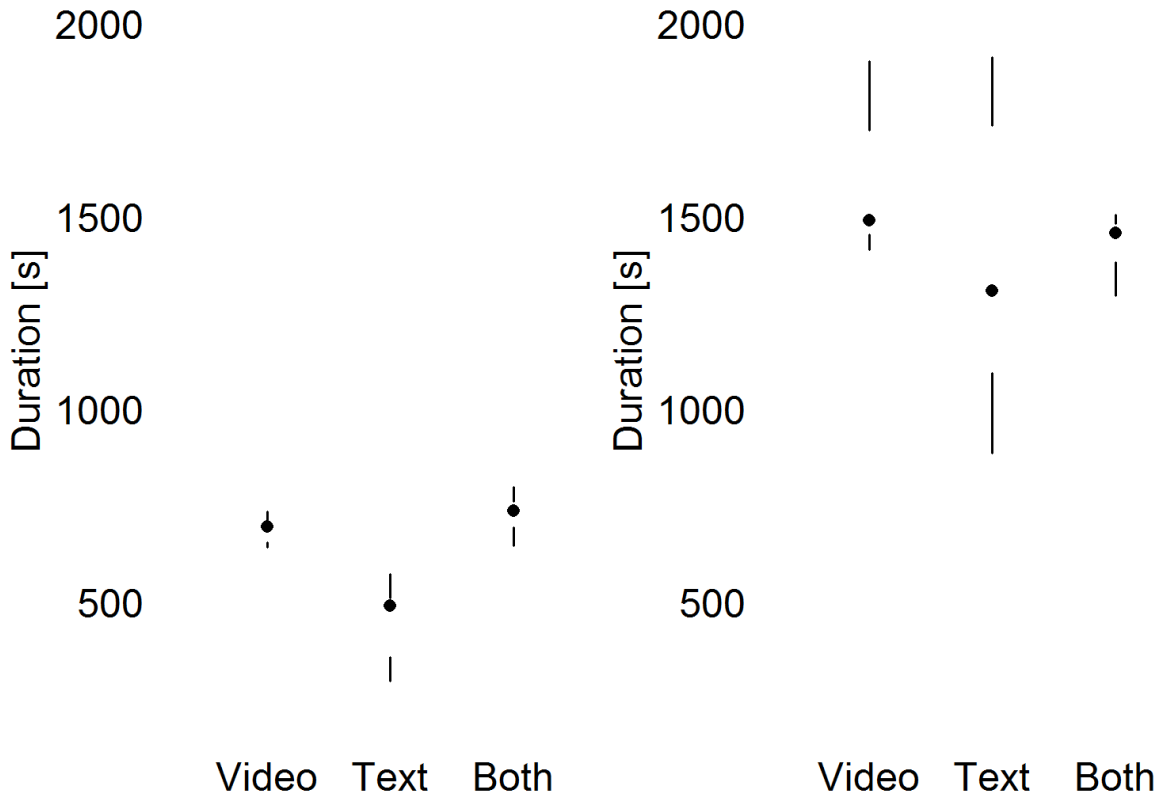
Fig. 9 shows how much the participants understood of the tutorials at the first attempt as assessed by themselves. There is nearly no difference between the text and the video tutorial. The values spread close to the mean ($\sigma_{Video}=18.321$, $\sigma_{Text}= 11.095$).

In Fig. 10 it can be seen how many correct and wrong answers the result files of the three tasks contained. The files from the first two tasks (Playground1 and Playground2) contained almost no errors. Conversely, the correctness of the files in the main task was very poor. As mentioned, we evaluated two unit tests for this task. As Fig. 10c shows, there were no significant differences between the three groups regarding the correctness of the first final task unit test. Also, when taking into account the second final task unit test the overall picture remains unchanged as shown in Fig. 10d.

Finally, Fig. 11 shows that there are small differences between the groups in how easy or hard they perceived the whole experiment to be. There is no clear tendency that one group had more or less challenges when solving the tasks in our study.

5.2 Data Set Preparation

To reconstruct the overwritten data mentioned in section 4.2, one of us watched the screen recording and in parallel repeated the clicks from the video.



(a) Duration of using the scenario tutorial in seconds (b) Duration of using the scenario tutorial and the following task in seconds

Figure 7: Duration of using the scenario tutorial and using the scenario tutorial plus the following task in seconds

Depending on the examined variables, we combined the results of some groups. For example, for the question which tutorial was used, we merged the two groups with both tutorials (C and F), no matter if with or without the scenarios as it was irrelevant in this case.

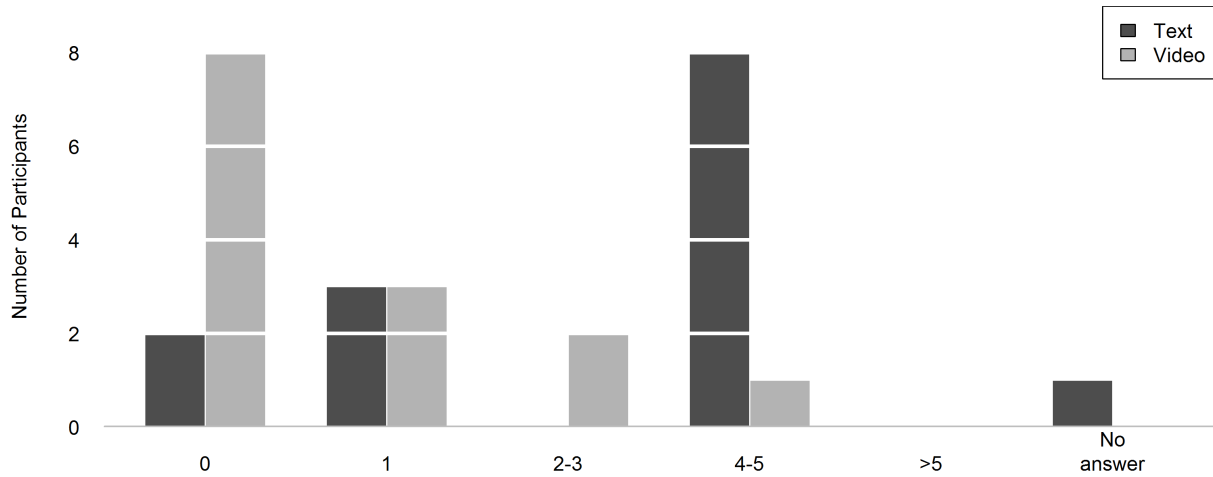
One problem was the question for groups C and F how many times they looked up information. Many used the option *Not used* as used zero times and not as the intended *I did not use this tutorial at all*. We found out about this because many participants stated that they used both tutorials equally in one question and then stated that they did not use one at all in the described question. We solved this problem by correcting the *Not used* answers to 0 times when the participants did not state *only video* or *only text* in the other question.

5.3 Hypothesis Testing

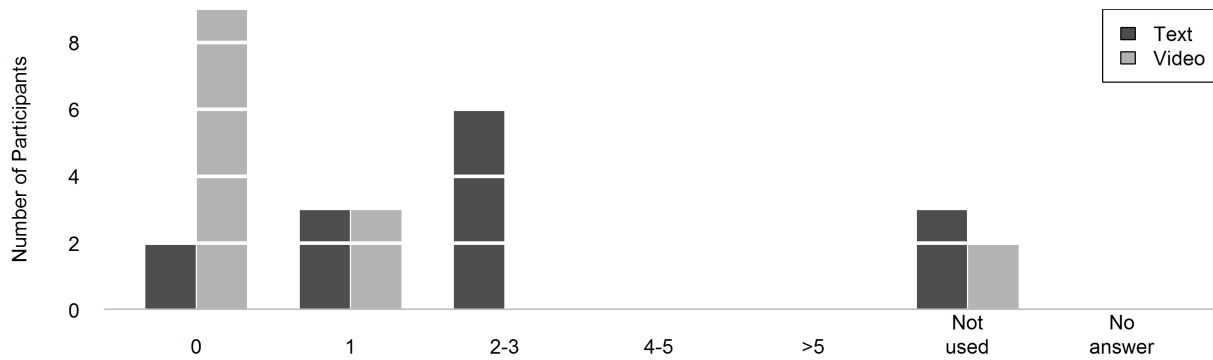
H_1 - Usage of the Tutorials The bar chart in Fig. 5 looks like there might be a tendency towards the video tutorial. We conducted a Shapiro-Wilk test and saw that the data is normally distributed ($p=0.065$) and there is no preference for videos. We cannot reject the null hypothesis.

H_{2a} - Duration of Tutorials As Fig. 6a shows, the *video and text* tutorial group took longer than the other two groups to complete the SIF tutorial. The text tutorial group was indeed significantly faster than the slowest *video and text* tutorial group with a strong effect size ($p_{Text-Both}=0.0257$, Cohen's $d=0.919$, $r=0.417$, $MD=152.407$). Between the other groups there was no significant difference ($p_{Video-Text}=0.65$, $p_{Video-Both}=0.0849$).

The results in Fig. 7a show that for the scenario tutorial, there were significant differences as well. This time the video tutorial took longer than the text tutorial with a very strong effect size ($p_{Video-Text}=0.0005$, Cohen's $d=3.14$, $r=0.843$, $MD=242.571$). Also, the *video and text* tutorial group took longer than the text tutorial group with a very strong effect size ($p_{Text-Both}=3.4722e-05$, Cohen's



(a) The participants with **only one** tutorial stated how many times they looked something up in the used tutorial



(b) The participants with **both** tutorials stated how many times they looked something up in which tutorial

Figure 8: How many times was an item looked up in which tutorial?

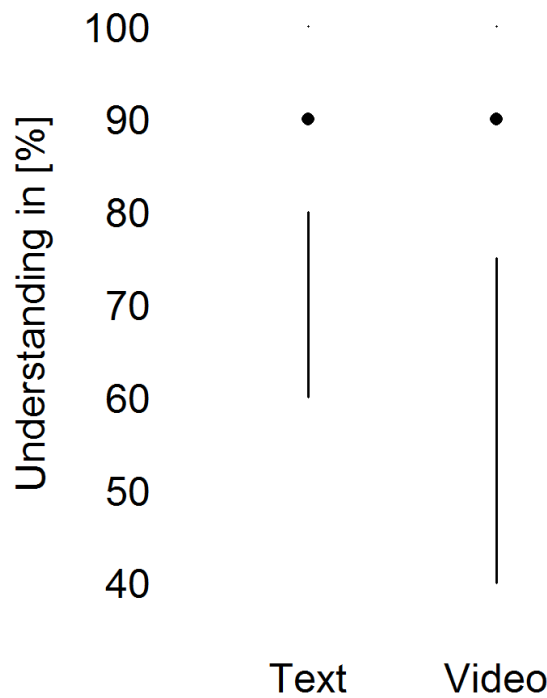
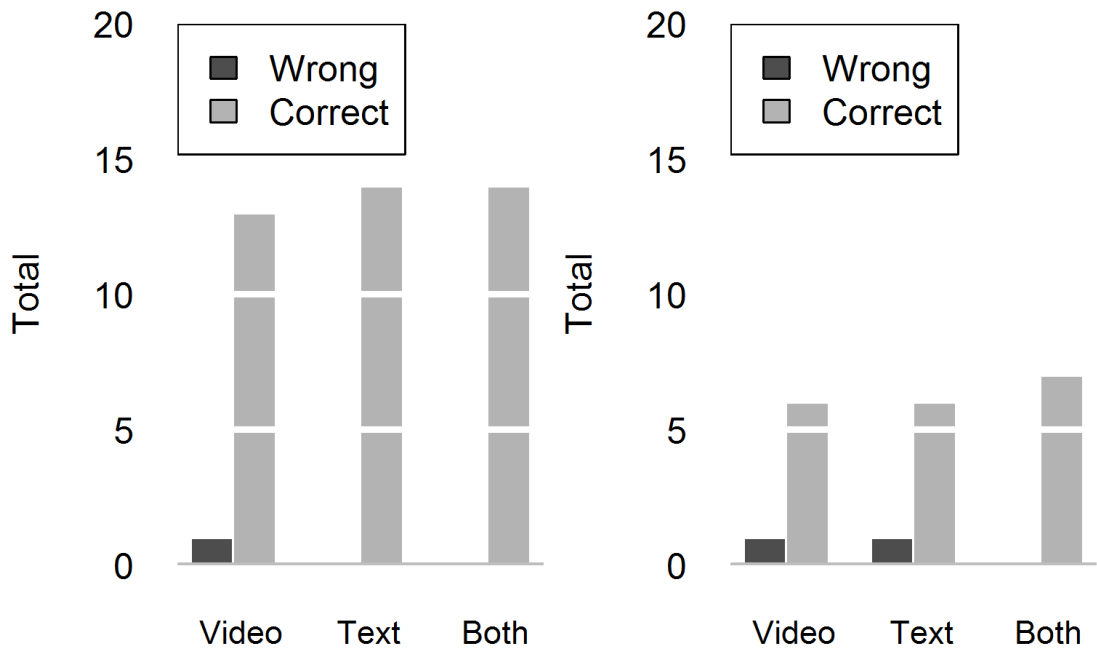
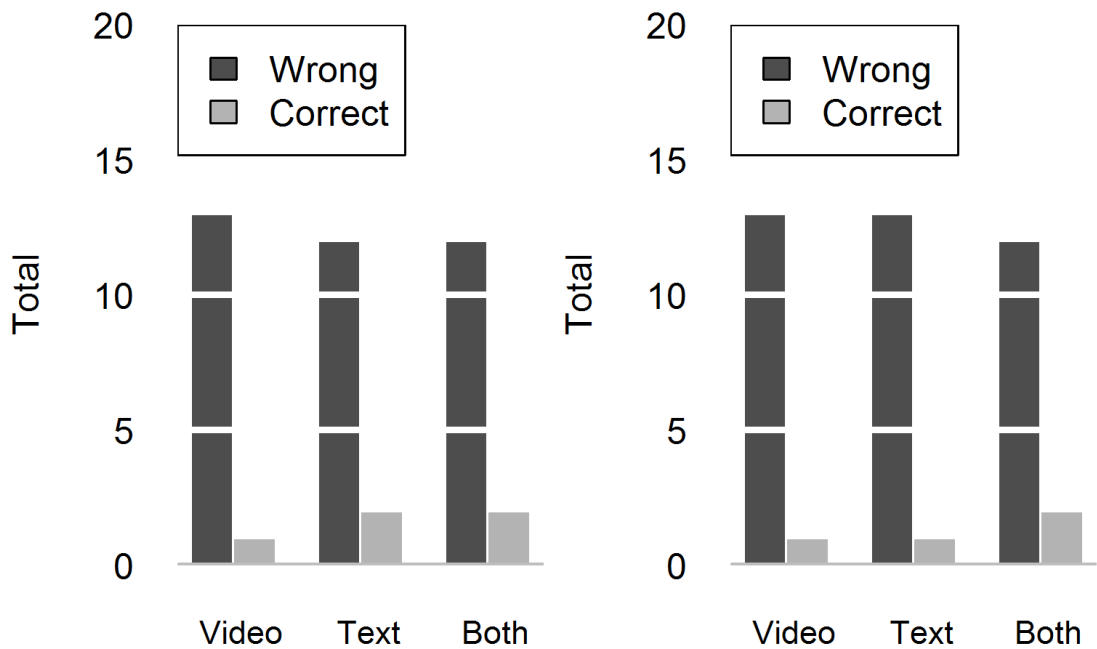


Figure 9: Q: What percentage of the tutorial did you understand when reading/watching it for the first time?



(a) Number of wrong and correct answers in the first task

(b) Number of wrong and correct answers in the second task



(c) Number of wrong and correct answers in the final task after inserting the new rate

(d) Number of wrong and correct answers in the final task after inserting the on-net texts

Figure 10: Number of wrong and correct answers in the tasks

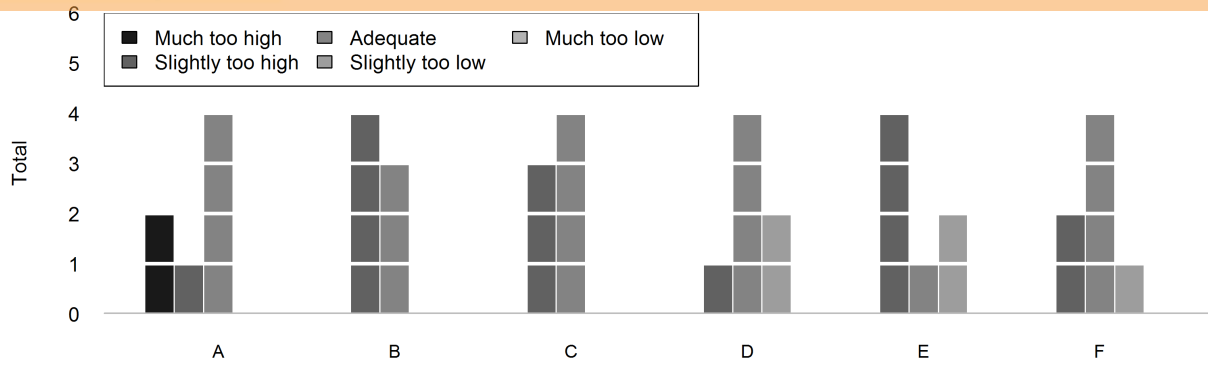


Figure 11: Difficulty level of our study perceived by the participants

$d=3.441$, $r=0.865$, $MD=285.286$). In this case there was no significant difference between the video tutorial group and the *video and text* group ($p_{Video-Both}=0.1058$).

The null hypothesis that all three groups need the same time can be rejected in both cases as there are differences between some of the groups.

H_{2b} - Duration of Tutorial and Task Concerning the time needed for the tutorial and the following task, the results are different. For the SIF tutorial and the task there are no significant differences between the three groups ($p_{Video-Text}=0.793$, $p_{Video-Both}=0.5112$, $p_{Text-Both}=0.6848$). The same result shows for the scenario tutorial and the following tasks ($p_{Video-Text}=0.3176$, $p_{Video-Both}=0.2593$, $p_{Text-Both}=0.724$). Therefore, the null hypothesis that there are no differences between the three groups cannot be rejected.

H_{3a} - Looking Things Up A test confirmed that the values in Fig. 8a are significantly different with a very strong effect size ($p=0.0067$, Cohen’s $d=1.274$, $r=0.547$, $MD=1.907$). The participants looked up more things in the text tutorial than in the video tutorial. The same applies for Fig. 8b with a very strong effect size ($p=0.0029$, Cohen’s $d=1.7$, $r=0.648$, $MD=1.15$). We can reject the null hypothesis that the participants look things up in both tutorials equally often. They looked up more items in the text tutorial.

H_{3b} - Understanding We thought that there is a difference in how much the participants understood when reading/watching the tutorial for the first time. With a p-value of 0.9462 we cannot confirm this result. We cannot reject the null hypothesis. There is no significant difference between the two tutorial types.

H_{3c} - Error Quota Regarding the number of correct and wrong answers for all four tasks (the final task was divided into two parts) we could not find any significant differences between the three groups. Table 1 shows the p-values of the significance tests. The null hypothesis cannot be rejected.

Table 1: P-values of the significance tests for comparing the results of the playgrounds and the final task

	Playground 1	Playground 2	Final task 1	Final task 2
Video-Text	0.3531	1	0.5774	0.6388
Video-Both	0.3531	0.3914	0.5774	0.6388
Text-Both	1	0.3914	1	1

H_{3d} - Level We checked with a t-test if there are significant differences in terms of the level between the three tutorial types or between the groups with the scenario task and the groups without it. Against our expectations, there are no significant differences ($p_{A-B}=1$, $p_{A-C}=0.7195$, $p_{B-C}=0.6589$, $p_{D-E}=0.3431$, $p_{D-F}=0.4744$, $p_{E-F}=0.6812$, $p_{A-D}=0.1023$, $p_{B-E}=0.7195$, $p_{C-F}=0.467$). We cannot reject the null hypothesis.

6 Discussion

6.1 Evaluation of Results and Implications

H_1 - Usage of the Tutorials Contrary to our hypothesis, there was no difference to be found regarding how often the different tutorial types were chosen. Apparently, the participants had no significant preference for one tutorial type. From what we could see during the experiments, many participants started with the video tutorial and then flipped through the text.

H_{2a} - Duration of Tutorial Usage For the SIF tutorial, the groups with both tutorials took longer to get through the tutorial. This might be because they had two sources and therefore had a look into both which, of course, takes longer.

There was no difference between text and video in this case, which might be because the whole program was new to the participants and, therefore, the participants with the text tutorial took their time to work through the text and fully understand everything.

For the scenario tutorial, again the group with both tutorials took longer than the text tutorial group, for the same reason. Additionally, for this tutorial the video group also took longer. This can have two reasons: either the text tutorial group was faster because they could easily skip uninteresting parts or because the scenario video took longer than the previous video and now the time difference was bigger.

All in all, we conjecture that the readers are faster because they can choose how fast they read. The video is always of the same length. The longer the video, the bigger the difference between video and text.

H_{2b} - Duration of Tutorial and Task The results of this hypothesis were quite unexpected. Although the text tutorial group was faster during the tutorial itself, they were not faster overall, including the following task. We have one explanation for this. As the participants looked up more things in the text tutorial, our conclusion is that the text tutorial group might not have understood as much as the other groups and therefore needed more time during the task to look up information. So the time saved due to fast reading is compensated for by looking up things more often.

H_{3a} - Looking Things Up Overall, most participants looked up more things in the text tutorial. This might be because it is easier to find things in a text than in a video. Maybe the video was better in explaining and, therefore, the participants did not have to look things up, though this is unlikely, as the participants stated to have understood both tutorial types equally well. Another explanation could be that the participants skipped some parts while reading. Both interpretations are possible.

H_{3b} - Understanding We first created the videos and then transcribed them into a text. That is why the given explanations are the same, just the format is different. This leads to the conclusion that it does not matter how an explanation is given, as long as the explanation itself is good. A bad explanation will not get better just because one uses a different format.

H_{3c} - Error Quota For the practicing tasks as well as for the final task all three groups made about the same number of errors. Contrary to our hypothesis, no tutorial taught the program better, which may have led to a better result in general. As the tutorials were equally good, there is also no difference in the result of the final task. This leads to the conclusion that the format of a tutorial is less important than the content. As the content was exactly the same, the results were the same as well. The many wrong results in the final task can be explained by the not that good Excel knowledge of many participants. The playgrounds show that it was possible to solve the tasks no matter which tutorial was used but for the more difficult final task the lack of Excel knowledge caused problems for some participants.

H_{3d} - Level Quite unexpected, having to go through the advanced tutorials with additional information to learn had no impact on how easy or difficult the participants perceived the experiment to be. We see two possible explanations here: either the format of the tutorial really makes no difference or the tasks and the technology under evaluation were too easy or too hard for a difference to become evident.

6.2 Threats to Validity

We have identified a number of threats to validity for our study and split them into three groups: construct validity (CV), internal validity (IV) and external validity (EV). We discuss them in the following:

- (CV) We took the opinion of the participants directly as a measurement. Still, it might be that some participants misjudged their usage of the tutorial. Nevertheless, this should balance out due to the number of participants.
- (CV) As the participants had the choice between three different experiments, the results cannot be mapped to computer science students from our university in general since probably students who are interested in Microsoft Excel actively decided to take our experiment.
- (CV) While the authors of the tutorials are experienced lecturers, they are not professional tutorial creators. Therefore, our tutorials might differ from those created by professional tutorial creators.
- (IV) The groups varied only by the kind of tutorial they received. Although we matched the participants and the tutorial groups completely randomly, the sample size might not have been sufficient to guarantee true randomness.
- (EV) Although we had a total of 42 participants, the sample size is not sufficient for generalizing our findings. Yet, we could see several statistically significant differences with medium to strong effect sizes.
- (EV) Instead of using experienced software developers we used software engineering students in our study. However, if even rather unexperienced students get along with a tutorial, we assume that the tutorial should be suitable for experienced professional developers as well.
- (EV) While for the purpose of our experiment it was productive to have a text tutorial which is equivalent to a video tutorial content-wise, in practice a combination with one tutorial type covering basic concepts supplemented by another tutorial type covering advanced topics might work even better.
- (EV) The study only investigated short-term learning effects, since the participants had to apply the learned content directly after consuming the tutorial. However, there could be significant differences when comparing long-term learning effects.
- (EV) The spreadsheets we provided to the participants contained seeded errors which is problematic as Panko explains [4]. While this might be regarded as a general threat for studies, we do not see a negative impact in the context of this concrete study.
- (EV) We only used content-equivalent representatives for two text and video tutorials which explain a single software tool and which were produced by the same authors. To further generalize our findings, it would be necessary to investigate more content-equivalent tutorials.

Overall, the answers we found seem reasonable to us. We confirmed them by the aforementioned statistical tests (t-test, Shapiro-Wilk test and the Wilcoxon signed-rank test) and found no contradictions. Nevertheless, we encourage other researchers to replicate our experiment to further confirm our results. SIF is available at the URL provided in [12]. A pre-print of this paper can be found at [11]. The raw data we used can be found at [14].

6.3 Lessons Learned

Regarding the experiment, we were surprised that so many computer science students had so little knowledge of Excel. This did not influence the perception of the tutorials but clearly influenced the many wrong results in the final task. For further experiments we would chose either selected participants with a broader knowledge of Excel or tasks which require less previous knowledge about Excel.

7 Conclusions and Future Work

7.1 Summary

In this study we wanted to investigate differences in educational effects between text and video tutorials. There was no clear preference of text or video tutorial. What we could see is that when provided with both, many participants first watched the video to get a general overview and then used the text to look things up. Regarding the time needed to complete the tutorials, we could see that the groups with both tutorials were needed more time than the other two groups and also that for the second longer scenario tutorial the video group was slower than the text group. There was no difference in how much the participants understood after reading/watching the tutorials for the first time, but if they had to look something up they used the text tutorial. We could not see any significant differences in the amount of right and wrong answers to the tasks which leads to the conclusion that both tutorial types taught the features of SIF equally well. There was also no difference in how easy or hard the participants thought the study was; therefore, we think that the amount of taught features does not influence the perception of the tutorials.

This leads to the conclusion that in an educational context it would be the best for software makers simply to provide developers with both text and video tutorials. As we assume that authoring text tutorials takes less effort than authoring comparable video tutorials, one could argue that text tutorials provide much better value. However, most developers prefer watching videos in the beginning instead of reading text. Therefore, software makers who only provide text tutorials will need to spend more effort in order to motivate developers to actually consume their text tutorial.

7.2 Future Work

There are many more different tutorial approaches than just text and video tutorials, e.g. interactive tutorials (though most of them are for games and fewer for application software). For further research it would be interesting to see how these tutorial types compare to our tutorials. Just because videos and text tutorials seem to be most prominent, this does not mean the lesser known and therefore lesser used methods have to be less effective.

Also, replicating our study with more participants or with typical end-users without a background or previous knowledge in software engineering could help to generalize our results. As it is commonly known, “people from IT” approach software applications differently from typical end-users. Therefore, it would be interesting to see how users from other fields work with unknown software, what problems they have and if they generally have a different approach towards tutorials than developers.

Another aspect might be the actual activity at which a tutorial type is targeted. Is it targeted just at teaching the basics or rather advanced topics?

As there were no significant differences regarding the results we think that it is indeed a personal choice what someone prefers. It would be interesting to see if other characteristics of a person influence this personal preference, e.g. the learning style or the knowledge of the language that is used. It might be easier to understand a video where one can see what is clicked and done than a written text if one is not that familiar with the used language or not that good at reading a long text (e.g. children).

References

- [1] Limesurvey. URL <https://www.limesurvey.org>.
- [2] R. URL <https://www.r-project.org/>.
- [3] K. Alexander. The usability of print and online video instructions. In *technical Communication Quarterly*, volume 22, pages 237–259, 2013.
- [4] S. Aurigemma and R. R. Panko. The Detection of Human Spreadsheet Errors by Humans versus Inspection (Auditing) Software. *Proceedings of EuSpRIG 2010 Conference Practical steps to protect organisations from out-of-control spreadsheets*, pages 1–14, 2010. URL <http://arxiv.org/abs/1009.2785>.
- [5] R. Baecker. Showing Instead of Telling. *Proceedings of the 20th annual international conference on Computer documentation*, pages 10–16, 2002. doi: 10.1145/584955.584957.
- [6] T. a. DeVaney. Impact of video tutorials in an online educational statistics course. *Journal of Online Learning and Teaching*, 5(4):600–608, 2009.
- [7] A. Dix, J. Finlay, G. D. Abowd, and R. Beale. *Human-Computer Interaction*. Prentice-Hall, Inc., 3 edition, 2004. URL http://fit.mta.edu.vn/files/DanhSach/_Human_computer_interaction.pdf.
- [8] T. Grossman, G. Fitzmaurice, and R. Attar. A Survey of Software Learnability: Metrics, Methodologies and Guidelines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 649–658. ACM, 2009. doi: 10.1145/1518701.1518803.
- [9] C. Ilioudi, M. N. Giannakos, and K. Chorianopoulos. Investigating differences among the commonly used video lecture styles. In *CEUR Workshop Proceedings*, volume 983, pages 21–26. CEUR-WS, 2013.
- [10] A. Jedlitschka, M. Ciolkowski, and D. Pfahl. Reporting Experiments in Software Engineering. In *Guide to Advanced Empirical Software Engineering*, pages 201–228. Springer London, 2008. doi: 10.1007/978-1-84800-044-5_8.
- [11] V. Käfer, D. Kulesz, and S. Wagner. What is the best way for developers to learn new software tools? an empirical comparison between a text and a video tutorial. 09 2016. URL <https://peerj.com/preprints/2413/>.
- [12] D. Kulesz. Spreadsheet inspection framework. URL <http://www.spreadsheet-inspection.org/index.html>.
- [13] D. Kulesz. From good practices to effective policies for preventing errors in spreadsheets. In *Proceedings of EuSpRIG 2011 Conference Spreadsheet Governance Policy and Practice*. EuSpRIG, 2011. URL <https://arxiv.org/abs/1111.6878>.
- [14] D. Kulesz and V. Käfer. Data for Spreadsheet Guardian, Nov. 2016. URL <https://doi.org/10.5281/zenodo.188896>.
- [15] D. Kulesz, J. Scheurich, and F. Beck. Integrating Anomaly Diagnosis Techniques into Spreadsheet Environments. In *2014 Second IEEE Working Conference on Software Visualization (VISSOFT)*, pages 11–19, 2014. doi: 10.1109/VISSOFT.2014.12.
- [16] D. Kulesz, F. Toth, and F. Beck. Live Inspection of Spreadsheets. In *Proceedings of the 2nd Workshop on Software Engineering Methods in Spreadsheets*, 2015. URL <http://arxiv.org/abs/1505.02428>.
- [17] D. Kulesz, V. Käfer, and S. Wagner. Spreadsheet guardian: An approach for protecting semantic correctness throughout the evolution of spreadsheets. submitted to arXiv, publication pending, temporarily available also from <ftp://ftp.informatik.uni-stuttgart.de/pub/se/plain.pdf>, 2016.
- [18] J. Lazar, A. Jones, and B. Shneiderman. Workplace user frustration with computers: an exploratory investigation of the causes and severity. *Behaviour & Information Technology*, 25(3):239–251, 2006. doi: 10.1080/01449290500196963.

- [19] S. A. Lloyd and C. L. Robertson. Screencast tutorials enhance student learning of statistics. *Teaching of Psychology*, 39(1):67–71, 2012. doi: 10.1177/0098628311430640.
- [20] A. P. Martin, M. Y. Ivory, R. Megraw, and B. Slabosky. Exploring the persistent problem of user assistance. Technical report is-tr-2005-08-01, Information School, University of Washington, August 2005. URL <https://digital.lib.washington.edu/researchworks/handle/1773/2079>.
- [21] L. S. Mestre. Student preference for tutorial design: a usability study. *Reference Services Review*, 40(2):258–276, 2012. doi: 10.1108/00907321211228318.
- [22] Microsoft. Excel 2013. Software.
- [23] S. Palmiter and J. Elkerton. Animated demonstrations for learning procedural computer-based tasks. *Hum.-Comput. Interact.*, 8(3):193–216, Sept. 1993. doi: 10.1207/s15327051hci0803_1.
- [24] S. J. Payne, L. Chesworth, and E. Hill. Animated demonstrations for exploratory learners. *Interacting with Computers*, 4(1):3 – 22, 1992. doi: [http://dx.doi.org/10.1016/0953-5438\(92\)90010-D](http://dx.doi.org/10.1016/0953-5438(92)90010-D).
- [25] C. Plaisant and B. Shneiderman. Show me! guidelines for recorded demonstration. In *IEEE Symposium on Visual Languages and Human-Centric Computing*, 2005.
- [26] E. R. Tufte. *The Visual Display of Quantitative Information*, volume 2. 2001. doi: 10.1198/tech.2002.s78.
- [27] D. H. van der Meij and D. J. van der Meij. Eight guidelines for the design of instructional videos for software training. *Technical communication*, 60(3):205 – 228, 2013.
- [28] H. van der Meij and J. van der Meij. A comparison of paper-based and video tutorials for software learning. *Computers & Education*, 78:150–159, 2014. doi: <http://dx.doi.org/10.1016/j.compedu.2014.06.003>.
- [29] H. van der Meij, J. Karreman, and M. Steehouder. Three decades of research and professional practice on software tutorials for novices. In *Technical Communication*, volume 53, pages 265–292, 2009.
- [30] B. van Loggem. 'Nobody reads the documentation': true or not? In *Proceedings of ISIC: the information behaviour conference*, 2014. URL <http://www.informationr.net/ir/19-4/isic/isic03.html#.V76f0eB8uUk>.

A Experimental Material

A.1 Used Tutorials

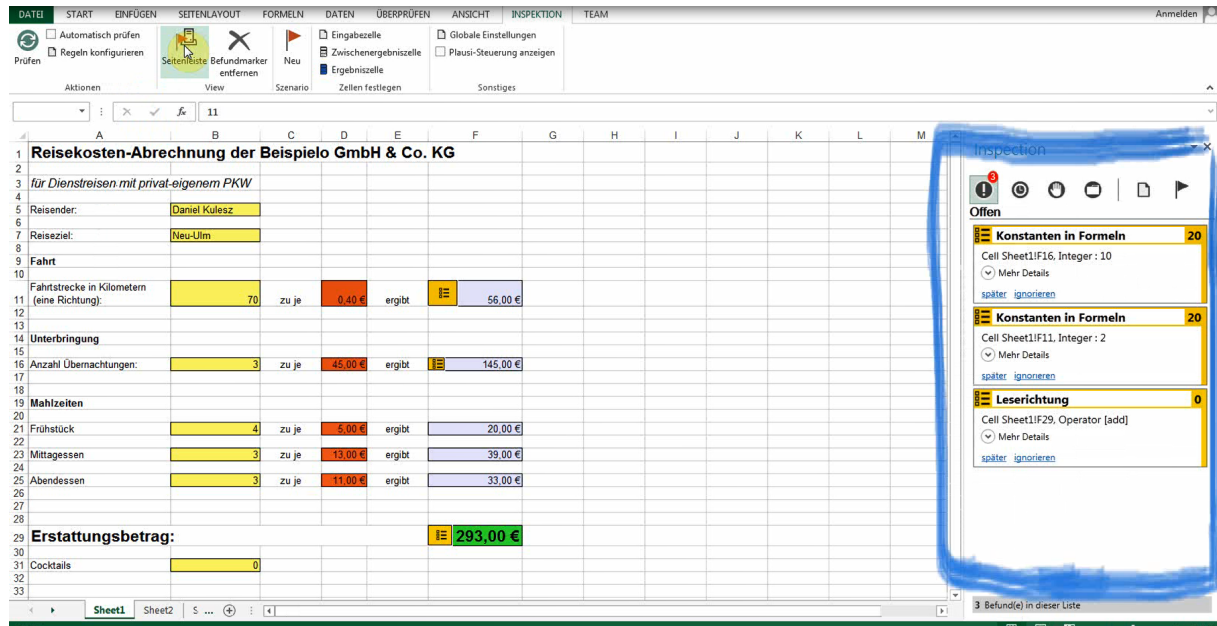


Figure 12: Screenshot of the video tutorial for SIF

4 The Side Bar

Using the button *Side Bar* (Seitenleiste) it is possible to fade in the side bar (Figure 10). The side bar list all issues among other things. The button for this is also placed in the inspection bar (Figure 9).

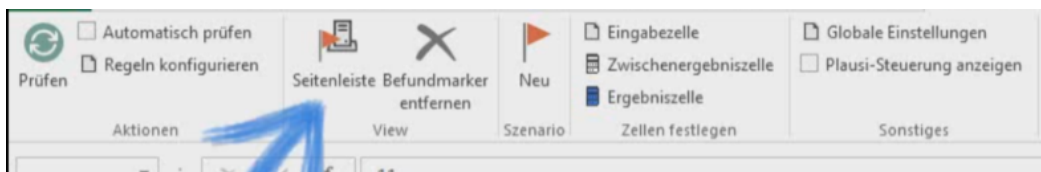


Figure 9: Button to fade in the side bar

A small number in a red circle informs how many issues have not been looked at yet. Like in an email program an issues is shown less prominent once it has been read.

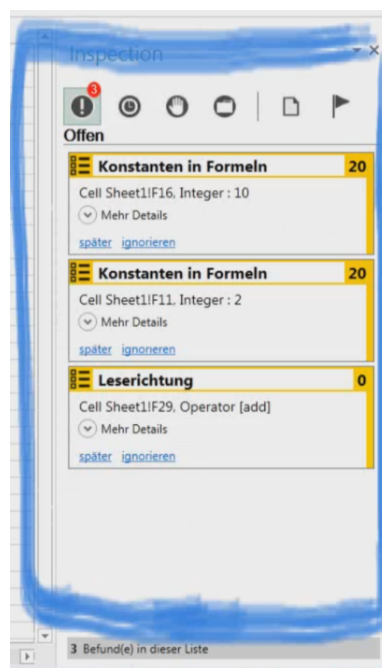


Figure 10: The side bar

Figure 13: Excerpt of the text tutorial for SIF (translated)

A.2 Pre-Test

How long have you worked with Excel?

For less than a year

For more than a year but less than four years

For more than four years

What do you use Excel for?

Data management (e. g. lists with names and addresses)

Tracking of data (e. g. exam performance, budgets, ledges, inventory ...)

Data analysis to find out trends and create forecasts (e. g. financial expenses, trend of numbers of course participants...)

Data analysis for decision support and selection between alternatives

Data analysis for scientific evaluations (e. g. experiments, experiment data ...)

Other:

As how important do you estimate the importance of Excel for your scope of work?

Unimportant

Rather unimportant

Quite important

Indispensable

How do you rate your Excel skills?

No or only little experience

Some experience but still a beginner

Plenty of experience but no expert yet

Expert

Which other programs apart from Excel do you use on a daily basis?

Email and internet (e. g. Microsoft Outlook, Mozilla Firefox, Microsoft Internet Explorer)

Office (e. g. Microsoft Word)

Simple databases (e. g. Microsoft Access)

Complex databases (e. g. MySQL, Microsoft SQL Server, Oracle Database, IBM db2...)

Other standard software (e. g. Maple, Catia, MS Project ...)

Programs for company management (e. g. SAP, Lexware ...)

Programs tailored to my special field

"When I typically work on Excel workbooks, I ...

... work on the Excel workbooks alone."

... work on the Excel workbooks in a team of 2-3 persons."

... work on the Excel workbooks in a bigger team with 4 or more persons."

If you get stuck while using Excel, where do you search for a solution **first**?

In the integrated help function

I ask co-workers or colleagues

In books/specialist literature

I look for a video on the internet

I read pages with tips on the internet

I ask a search engine about the problem and check out what it gives

Other:

Which additional aids for your work with Excel have you used to date?

None, I made it through on my own.

The integrated help function

(Small) courses or workshops

Occasional help from colleagues and co-workers

Figure 14: Pre-test, page 1 (translated)

Pages with tips on the internet
 Books and/or other specialist literature
 Which share of your working time do you spend with Excel?
 0-25%
 26-50%
 51-75%
 76-100%
 With how many workbooks do you typically work per week?
 0-1
 2-5
 6-10
 More than 10
 I can't estimate.
 Which techniques do you typically apply to compare different versions of your workbooks? (versioning)
 None, for each of my workbooks there is normally only the last "current", version
 I save workbooks with the current date and/or the current time.
 When saving, I give the workbooks a version number.
 I save the workbook with my username or short name.
 Other:

 Do you sometimes use workbooks that you did not create yourself? (but someone else)?
 No, I always use workbooks which I created myself.
 Yes, I sometimes use workbooks which I did NOT create myself.

Do you sometimes also create or change formulas in workbooks?
 No, I only enter data - if at all.
 Yes, I create or change formulas.
 How often do you create or supplement documentation for workbooks you work with (e. g. comments in cells, explanations etc.)?
 Practically never
 Rarely
 Often
 Very often
 What do you typically do if a word in a workbook looks strange to you?
 I don't do anything but simple ignore that.
 I ask someone who knows more about it before trying to comprehend the value.
 I first try to figure out myself how the value is calculated.
 Do you regularly share the results of your work with Excel (e. g. the completed address list, the updated balance ...) with other persons?
Note: This question tackles sharing of the results in general, no matter if you share the the whole workbook or not.
 No, I use the results only myself.
 Yes, but only in few cases.
 Yes, in most cases.
 (Almost) always. Actually, that's the reason why I create these results at all.

How may we contact you?
We will use this data only for contacting you. We will not share or publish this data under any circumstances.
 Name :

 Email:

 Phone:

Figure 15: Pre-test, page 2 (translated)

A.3 Final Test

<p><i>Final test for the Excel study 06/2016 (C)</i> <i>D. Kulesz, V. Käfer, ISTE / Uni Stuttgart</i> 1</p>
<h2 style="margin: 0;">Final test for the Excel study 06/2016</h2> <p style="margin: 10px 0 0 0;">For the questions with tick boxes, please only tick one answer! <i>(If not stated otherwise)</i></p>
<p>Meta</p> <p>----- <i>I will fill out this section (questions 1-3) for you!</i> -----</p>
<p>1. Registration number:</p> <p>_____</p>
<p>2. Experiment:</p> <p>_____</p>
<p>3. Activities according to first survey:</p> <ul style="list-style-type: none"> <input type="checkbox"/> Documentation <input type="checkbox"/> Development <input type="checkbox"/> Inspection <input type="checkbox"/> Sharing
<p>General</p> <p>----- <i>From here, now it's your turn...</i> -----</p>
<p>4. How did you like the study overall?</p> <ul style="list-style-type: none"> <input type="checkbox"/> Very good <input type="checkbox"/> Good <input type="checkbox"/> Not so good <input type="checkbox"/> Not good at all
<p>5. The difficult level for me was</p> <ul style="list-style-type: none"> <input type="checkbox"/> Much too high (I was overstrained) <input type="checkbox"/> Slightly too high (it was quite challenging) <input type="checkbox"/> Adequate <input type="checkbox"/> Slightly too low (it was a bit too easy) <input type="checkbox"/> Much too low (I was bored)
<p>Spreadsheet Inspection Framework</p>
<p>6. How did you like our tool 'Spreadsheet Inspection Framework' overall?</p> <ul style="list-style-type: none"> <input type="checkbox"/> Very good <input type="checkbox"/> Good <input type="checkbox"/> Not so good

Figure 16: Final test, page 1 (translated)

Please indicate how strongly you agree with the following statements:

7. I would also use the Spreadsheet Inspection Framework for my own workbooks.

Strongly disagree Strongly agree

8. I am confident that the workbook 'tarifvergleich.xlsx' still gives correct output values even after my modifications.

Strongly disagree Strongly agree

9. The optical design of the Excel workbook 'tarifvergleich.xlsx' is well done.

Strongly disagree Strongly agree

10. The maintainability of the Excel workbook 'tarifvergleich.xlsx' is good.

Strongly disagree Strongly agree

11. The existence of the scenarios has eased the maintenance of the Excel workbook 'tarifvergleich.xls' for me.

Strongly disagree Strongly agree

12. I would also check my Excel workbooks for compliance with the rule 'reading direction'.

Strongly disagree Strongly agree

13. I would also check my Excel workbooks for compliance with the rule 'constants in formulas direction'.

Strongly disagree Strongly agree

14. I would also check my Excel workbooks for compliance with the rule 'same references in a row'.

Strongly disagree Strongly agree

15. I would also check/safeguard my Excel workbooks using the scenario technique.

Strongly disagree Strongly agree

16. What did you like about the 'Spreadsheet Inspection Framework'? Please name a few concrete items, e.g. 'fast execution of the inspections':

1. positive item: _____

2. positive item: _____

3. positive item: _____

4. positive item: _____

17. In your opinion, what could be improved in the 'Spreadsheet Inspection Framework'? Please give a few short keywords, e. g. 'nicer colors for the icons':

1. improvement suggestion: _____

2. improvement suggestion: _____

3. improvement suggestion: _____

4. improvement suggestion: _____

Figure 17: Final test, page 2 (translated)

Tutorial

18. Did you look up a function you were not familiar with?

- No
- Yes, using the following approach (e. g. video, test on website):

19. How often did you look up something in the video (estimate)?

- _____ times
- Not used

20. How often did you look up something in the text tutorial (estimate)?

- _____ times
- Not used

21. Which tutorial did you use more often?

Only video ------- Only text

22. What did you like about the video tutorials? Please name a few concrete items, e.g. 'good explanations':

1. Positive item: _____
2. positive item: _____
3. positive item: _____
4. positive item: _____

23. What did you dislike about the video tutorials?

1. negative item: _____
2. negative item: _____
3. negative item: _____
4. negative item: _____

24. What did you like about the text tutorials? Please name a few concrete items, e.g. 'good explanations':

1. positive item: _____
2. positive item: _____
3. positive item: _____
4. positive item: _____

25. What did you dislike about the text tutorials?

1. negative item: _____
2. negative item: _____
3. negative item: _____
4. negative item: _____

Figure 18: Final test, page 3 (translated)

26. What percentage of the videos did you approximately understand when watching them for the first time?

_____ percent

27. What percentage of the text tutorials did you approximately understand when reading them for the first time ?

_____ percent

Activities

 Please try to answer the following questions regarding your habits / activities from the perspective you had **before** participating in this study!

Activity: Documentation

28. Which techniques do you apply to document your workbooks?

Several answers allowed - please choose at least one answer.

- I put small descriptive texts directly into the affected worksheets.
- I put comments in cells.
- I create a separate worksheet for the documentation.
- I create a separate document (e. g. a Word file) for the documentation.
- Other: _____

Activity: Development

29. When you create new workbooks, do you start with a completely new, empty workbook?

- Yes, (almost) always.
- Sometimes yes but sometimes I use an existing workbook as a foundation.
- No, I (almost) always rely on existing workbooks.

30. What is typically your first step when creating a new workbook?

- I take a look on other, similar workbooks to gather some design ideas.
- I design the structure of the new workbook with pen and paper.
- I enter the most important formulas into the new workbook first.
- I start directly by entering the actual data.
- Other: _____

31. How big do the workbooks you create get during use later ? Note: By cells we mean all non-empty cells - that is, all cells which contain data or formulas.

- Less than 100 cells
- 101 to 1000 cells
- 1001 to 10000 cells
- More than 10.000 cells
- Other average value: _____
- Strongly depends on the particular workbook, giving an average value does not make sense
- I can't estimate.

Figure 19: Final test, page 4 (translated)

32. How long does a workbook created by you live on average, before it gets abandoned or deleted?

- About a week
- Several weeks or months
- One or more years
- Other average value: _____
- Strongly depends on the particular workbook, giving an average value does not make sense
- I can't estimate

33. Do you separate your workbooks in modules? (e. g. by segmenting inside one worksheet or by using multiple worksheets)

- No, never.
- Yes, but only occasionally
- Yes, most times
- Yes, (practically) always.

Activity: Inspection

34. Which of the following triggers make you typically inspect a workbook?

Several answers allowed - please choose at least one answer.

- I notice something strange, e. g. an implausible value in a cell.
- I have made changes to my workbook.
- When I work with workbooks I inspect them by principle.
- I am about to share a workbook or the results calculated by it.
- I receive a workbook on which someone else has worked.
- Other: _____
- I don't know.

35. What percentage of the time that you work with your workbooks do you typically spend for their inspection (e. g. checking if the workbook works correctly or systematically searching for faults)?

- 0-10%
- 11-25%
- 26-40%
- More than 41%
- I can't estimate.

Figure 20: Final test, page 5 (translated)

36. Which of the following techniques do you apply for the inspection of a workbook?

Several answers allowed - please choose at least one answer.

- Flying over the workbook with the eyes ('Eyeballing')
- Function 'Show all formulas'
- Manual inspection of specific cells/formulas
- Computation of selected cells with a calculator
- Function 'Edit - go to - contents'
- Adhoc-test with input values which just cross my mind
- Specially crafted inspection tools (e. g. 'Spreadsheet Detective')
- Systematic test of all formulas with previously planned input values and comparing expected and actual output values
- Thorough inspection with multiple reviewers, review assignment and preparation
- Common sense
- Other: _____

Activity: Sharing**37. How do you typically share the results you achieved with Excel (e. g. a completed address list, an updated balance)?**

Several answers allowed - please choose at least one answer.

- I print the workbook or parts of it and share the printout.
- I only share a summary of the relevant results.
- I share selected parts of the workbook.
- I share an export (e. g. PDF) of the complete workbook.
- I share the complete workbook as a whole
- other: _____

38. How often does this sharing occur?

- (almost) daily
- Once per week
- Once per month
- Once per quarter
- Less often than once per quarter

39. Which of the following techniques do you apply on workbooks which you share in order to protect them from unwanted changes?

Several answers allowed - please choose at least one answer.

- I mark areas explicitly (e. g. input values, output values) in order to save other users from errors.
- I protect the whole workbook with a password so that it is only readable.
- I protect certain cells / areas with a password.
- I don't have to do anything as the workbooks have already been protected by the 'previous owner'.
- I don't have to do anything because I don't share the workbook as a whole.
- None - I just hope that nothing will break even without safety measures.
- Other: _____

Figure 21: Final test, page 6 (translated)

Conclusions from the study

40. Now, that is – after participating in this study, would you change something regarding your habits?

- No
- Yes, the following (please give only short keywords):

Data regarding your person

41. Your gender:

- Female
- Male

42. To which age group do you belong?

- Under 20
- 20-25
- 26-30
- 31-35
- 36-40
- Above 40

43. What amount of 'full time work experience' do you have? (All times in which you earned 'real' money account to this, i. e. without times of training, studies, part time jobs or internships)

- Practically none yet
- Up to one year
- 1-2 years
- 2-5 years
- 5-10 years
- 10-20 years
- Over 20 years

Figure 22: Final test, page 7 (translated)

44. In which sector are you active resp. do you see yourself in?

Several answers allowed - please choose at least one answer.

- Agencies, promotion, marketing und PR
- Banking
- Building industry
- Mining
- Education and training
- Chemical and oil processing industries
- Printing, paper and packaging industries
- Electro technology, fine mechanics and optics
- Energy, water supply and disposal
- Automotive
- Financial service providers
- Leisure, tourism, culture and sports
- Health and social services
- Glass and ceramics
- Wholesale and retail
- Wood and furniture industries
- Hotel, gastronomy and catering
- Real estate
- IT and Internet
- Consumer goods
- Aviation and space travel
- Mechanical engineering and plant construction
- Media
- Medical technology
- Metalworking industry
- Food and luxury foods
- Civil service and federations
- Human resources
- Pharmaceutical industry
- Telecommunications
- Textiles, clothing and leather goods
- Transportation and logistics
- Consulting, auditing and law
- Insurances
- Science and research
- Other: _____

45. Do you have any comments regarding this questionnaire?

— Thank you very much for your participation! —

Figure 23: Final test, page 8 (translated)