"I did use it!" - Assessing subjective vs objective cognitive artifact usage

Stefan Morana[1], Julia Kroenung[2], and Alexander Maedche[1]

[1] Karlsruhe Institute of Technology, Institute of Information Systems and Marketing (IISM) and Karlsruhe Service Research Institute (KSRI), Karlsruhe, Germany
{stefan.morana,alexander.maedche}@kit.edu

[2] University of Mannheim, Dieter Schwarz Foundation Endowed Chair for E-Business and E-Government, Mannheim, Germany
{kroenung}@bwl.uni-mannheim.de

Abstract. For decades, IS researchers have discussed the reliability of subjective measurements to assess actual artifact usage. Especially in experimental settings, as in the design science context for instance, the participants' usage data of the evaluated artifact represents an important point of analysis. However, collecting objective usage data, (i.e. logfiles) is often not feasible depending on the artifact. In this paper, we present the theoretical grounding of collecting cognitive artifact usage data using eye-tracking technology. Grounded in immediacy and eye-mind assumption the participants' artifact fixations are used as objective usage measurement. The question remains if in comparison, the collection of subjective (e.g. perceptual) usage data is sufficient and reliable for such experiments. The results of our comparative analysis indicate that researchers could use subjective measurements when comparing different artifact designs and should rely on objective measurements when testing the effect of an artifact compared to a control group without artifacts.

Keywords: Artifact, eye-tracking technology, laboratory experiments, objective usage, subjective usage.

1 Introduction

The use of information technology (IT) by individuals remains the variable of interest in the information systems (IS) field and has been recognized as a key element of the *missing link* from IT investments to business value [1, 2]. The question if a certain IT artifact is used or not and which outcomes are related to the IT usage, has been subject of investigation in numerous experiments in IS research. In general, there are two possible approaches for measuring usage: subjective or objective measurements [3]. Already in 1989, Fred Davis concluded that "n*ot enough is currently known about how accurately self-reports reflect actual behavior*" [4: p. 334] and still, more than 25 years later, there is an ongoing debate on the reliability of self-reported, subjective usage measurements.

13th International Conference on Wirtschaftsinformatik,
February 12-15, 2017, St. Gallen, Switzerland

Morana, S.; Kroenung, J.; Maedche, A. (2017): "I did use it!" - Assessing subjective vs objective cognitive artifact usage, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 1021-1035

Overall, there is a strong tendency that subjective data is no fully reliable approximation of objective data in the research discourse [5]. Still, there are inconsistent results that argue for the reliability of subjective usage [6, 7].

The primary objective of experiments in the design science research (DSR) [8] context is the investigation if certain effects occur while participants use an IT artifact (we use the term artifact to subsume the terms artifact / treatment / intervention) [9]. In DSR, the usage of the evaluated artifact in the experiment is an important information that requires special attention [10]. Within an experimental setting, researchers can collect subjective data (e.g. by asking the participants on their artifact's usage in a questionnaire) or objective data (e.g. by recording the participants' interaction with the artifact using a screen cam). In most experimental settings, the participants actively use the artifact, from a human-computer interaction (HCI) point of view, for example, by clicking on a button or scrolling on a website. In such experimental settings, assessing the participants' objective usage is easy. For an overview on objective use measurements in experiments see Eckhardt et al. [3].

In contrast to this, there are experimental settings in which the participants do not actively use the artifact (e.g. clicking a button) during the experiment, but use it from a cognitive point of view. One example is the provision of a textual explanation [11] (as the actual artifact) in an experimental application. The participants can read these explanations during the experiment and use the provided textual information without any active interactions with the computer. Rather, the participants read the information, store it in their memory, and use it, for example, to process the experimental task. We define this as cognitive artifact usage; the user interacts with the artifact only by viewing at it and processing the provided information, for example, text or images, cognitively. Due to missing direct human-computer interactions with the artifact (e.g. by counting click rates), the logging of such objective usage data is more difficult. However, in most cases, it is important to know if and to which extent the participants used the artifact during the experiment to assess its effect or impact. In such experimental settings, researchers often rely on subjective usage data, because objectively measuring such cognitive artifact usage is difficult.

A possibility to address this shortcoming is the utilization of eye-tracking technology. Eye-tracking technology enables researchers to measure the participants' eye movements on the screen and thus, allow an objective measurement of the users' cognitive artifact usage. According to the immediacy assumption and the eye-mind assumption [12], researchers can infer the participants' current cognitive activity based on their current glance on the computer. The following example shall illustrate this assumption: While looking at a picture in a gallery, the individual most likely also thinks about the picture and thus, cognitive processes take place. Consequently, eye-tracking technology enables researchers to derive the participants' current cognitive activity by recording their eye fixations during the experiment. However, the utilization of eye-tracking technology in experiments today is still rather expensive from both, a financial and an effort point of view, compared to gathering subjective data using a survey. Thus, gathering subjective usage data is less effort. Considering the ongoing debate on subjective vs. objective measurement, the question remains if subjective measurement data is reliable for assessing cognitive artifact usage in laboratory experiments as formulated:

Do subjective measurements reliably approximate objective measurements with respect to cognitive artifact usage in laboratory experiments?

Our research contributes to the ongoing debate if subjective usage measurements are a reliable approximation for objective usage measurements in laboratory experiments. We provide insights from our DSR laboratory experiment and discuss under which conditions the application of subjective measurements is sufficient and when objective measurements are required based on our experiment data. From a methodological point of view, we present the theoretical baseline for objectively measuring cognitive artifact usage in laboratory experiments with eye-tracking technology, grounded in the immediacy assumption and the eye-mind assumption [12]. We demonstrate how to apply eye-tracking technology for measuring and analyzing objective artifact usage. In addition, our findings contribute to the IT usage research stream as they provide an example under which conditions subjective usage measurements can be applied.

2 Theoretical Foundations and Related Work

2.1 Information Systems

Researchers conduct experiments to evaluate if the usage of a certain artifact results in the proposed effect or not. There are two types of usage measurements within experimental setting, participants report their subjective usage, or researchers measure the usage objectively during the experiment. Wu and Du [13], based on a comprehensive literature review on IS usage research, discuss two different types of subjective usage types: reported usage and assessed usage. The researchers argue that the key difference between reported usage and assessed usage is the application of ordinal scales for the measurement. Reported usage indicates the "*users-reported amount of time or number of times of using a system*." [13: p. 683] In contrast to this, assessed usage "*refers to the ordinal-scale-measured intensity and extent of using a system*." [13: p. 683] For the remainder of the paper, we refer to subjective usage for both type of usage types.

Using subjective (reported and assessed) usage measurements can result in non-reliable data and distorted results, due to possible common method biases [7, 14]. There is an ongoing debate whether subjective usage is a reliable approximation for objective usage. When talking about artifact usage in experiments, the whole stream on IS usage research is important to consider. Before summarizing a selection of findings on the subjective vs. objective usage debate, we shortly address the construct system usage itself. In decades of IS usage research there is few "*in-depth, theoretical assessment of the construct*" [15: p. 228], until it was re-conceptualized by Burton-Jones and Straub [15]. The researchers propose six measures for system usage ranging from very lean to very rich, enabling researchers to provide more insights in the applied usage construct. Straub et al. [7] also discuss the assessment of system usage and the comparability of subjective and objective system usage. They found, in contrast to their expectations, that both types of usage measurements are not strongly correlated to each other [7].

In a repetition of the study by Straub et al. [7], Barnett et al. [6] found a stronger relationship between the subjective and objective system usage. Their analysis reveals a correlation between the perceived and objective usage, ranging from $r = 0.38$ to 0.43 [6]. The researchers conclude their article with a suggestion to consider the context of the study when deciding which type of usage measurements to apply. Barnett et al. [6] suggest to apply objective data when the actual usage is connected to the dependent variable (such as performance) and subjective data when general insights are required [6: p. 81]. In the discussion of their empirical results, Burton-Jones and Straub [15] also shortly address the question when to use subjective and objective measures. They suggest objective usage "*to measure the system and task aspects of usage*" and subjective usage (such as questionnaires) "*to measure user states such as cognitions or emotions during usage*". [15: p. 243]

2.2 Psychology

Investigating the discrepancy between self-reported and actual behavior has a tradition in experimental psychology and is related to the debate on usage behavior in IS research. There is evidence that participants are unaware of many biases in their choices resulting in questioning the reliability of subjective behavior compared to objective behavior [16]. One stream of research in this discipline further addresses introspective information and the reliability of such subjective data for decades. The reliability of subjective data measurements depends on the applied methods. Verbal reports for instance, can be accurate and a significant reflection of the objective behavior, if the introspective information is relevant to the current task and available in the participants' short-term memory [17, 18]. Thus, introspection can only access cognitive representation kept in the participants' working memory [17]. Other information influences the participants' introspective information. Studies revealed, the participants subjective measurements can be influenced by the perceived difficulty of the task [19] or subjects' own motor responses [20].

2.3 Human-Computer Interaction

In contrast to these supportive findings on a correlation between subjective and objective usage data in psychology, there are also non-supportive results. Analyzing participants' mobile application usage, Reuver and Bouwman [5] found no correlation between subjective and objective usage measurements. Moreover, they discuss the possibility of type I and type II errors when relying on subjective measurements alone [5]. All in common, in the aforementioned usage studies is the participants' active interaction with the system, for example, Straub et al. [7] investigate voice mail usage and Reuver and Bouwman [5] investigate mobile applications use. In such settings, the objective artifact's usage measurement is possible, for example, by using logging functionality within the applications.

This is not possible for assessing the artifacts' usage without some sort of direct human-computer interaction, for example, clicking on a certain button. An example of such a setting is the provision of an explanation within the experiment's application, which is always present and there is no activity by the participants required to make the explanation visible. The participant can read the explanation or not, there is no possibility to log this reading process within the experiment's application. The participant is looking at the explanation, reading it, and cognitively processing the provided information. Eye-tracking technology enables researchers to capture this reading or looking process by recording the participants' eye movements on the screen [21, 22]. This is grounded in the immediacy assumption and the eye-mind assumption proposed by Just and Carpenter [12, 22]. The immediacy assumption "*is that a reader tries to interpret each content word of a text as it is encountered.*" [12: p. 330] The eye-mind assumption "*is that the eye remains fixated on a word as long as the word is being processed.*" [12: p. 330]

Eye-tracking technology is also used for quite some years in HCI research [23] and psychological research [21, 24–26]. Several studies showed the linkage between the humans' eye movements and their attention [22, 24]. With respect to this research, we focus on this very link between eye fixations and attention, since the link between fixations and cognitive processing lies beyond the scope of this article. Although research in the field of cognitive psychology regards eye-fixations also as approximation for cognitive processing [24]. Eye-tracking technology is, for example, used to investigate eye movements in reading and information processing research [24, 25]. In a recent study, the participants' subjective report of eye fixations was compared to the objectively measured fixations using eye-tracking technology [27]. In this study, the participants were able to report at least a subset of their eye movements. Nevertheless, the researchers also identified false reported eye movements [27]. In line with the results on subjective vs. objective usage measurements [6], there is the potential for subjective measurements to be a reliable approximation for objective measurements, but it depends on the study context and applied methods. The eye-tracking technology provides researchers "*quantitative and qualitative measures of observers' subjective reports and reveal experimental effects of visual search that would otherwise be inaccessible." [27: p. 1]* Depending on the experimental setting, participants might not be able to provide a justification for their own decisions [28] and thus, eye-tracking technology can be used to obtain objective data [21].

3 Hypothesis development

Summarizing the brief presentation of the ongoing discussion on subjective vs. objective usage measurement, there are findings supporting a correlation [6, 7] and findings rejecting a correlation [5]. In line with the supportive findings, we assume that a reliable rating of the participants' cognitive artifact usage in a laboratory experiment is possible. In such a setting, the participants use the artifact cognitively, for example, by reading or viewing at it, based on the eye-mind assumption [12]. We assume these cognitive processes are stored within their short-term memory. According to the immediacy assumption [12], participants should be able to recall their recently conducted activities. Thus, when asking the participants immediately after the experiment, their activities during the course of the experiment should be available in their short-term memory. This relationship is additionally grounded in research on introspective information [17, 18]. Thus, we argue the participants' perception of their artifact usage should be possible. We propose that the subjective artifact usage is a reliable measurement in experimental settings with cognitive artifact use. Summarizing our assumptions, we propose the following hypothesis:

Hypothesis*: Subjective measurements are positively related to objective measurements of cognitive artifact usage in laboratory experiments.*

For testing this hypothesis, we consider the objective artifact usage as the reliable measurement. We test if there is a significant positive relationship between the subjective perceptions of artifact usage and its objective measurements. This would indicate that the perceived usage measurements are a reliable approximation for the objective usage measurements as tested in several other studies [5–7]. In order to prevent possible type I and II errors [5], we also test for a relationship between the perceived artifact's usage and the objective usage of other applications used in the experiment. This test is required in order to ensure that the participants' subjective measurement, their introspective information, only reflects the artifact's usage and not the usage of other applications. Next, we describe the experimental setup for testing our hypothesis.

4 Experimental setup

We test our hypothesis in the context of an experiment on the effect of providing textual and visual explanations [11] to support individuals' process execution. The experiment itself is part of a DSR project aiming to design an assistance system [29] to support users' process knowledge and process execution performance [30]. In the experiment, we simulated an IT ticketing process and tested if the explanations' provision by the assistance system has the intended effects on the users. The participants had to handle eight IT requests (e.g. create a new user account or purchase of hardware) provided by a simulated email client in a simulated ticketing system according to the specified ticketing process. This ticketing process is a simplified version of a real-world organizational ticketing process adapted from the DSR project's case company. The applications used in the experiment (see **Figure 2**) are simplified versions of the case company's applications used to handle their ticketing processes. One week before the experiment, the participants received a training on the experimental ticketing process and applications. They also received the documentation and we asked them to study it in detail.

The participants in the experiment are students (bachelor and master level) from a large German university and we randomly assigned them to one of three groups. We developed two versions of the assistance system as treatments. In both versions, the assistance system is depicting the sequence of the ticketing process steps supporting the participants with its execution. Moreover, the "advanced treatment" additionally provides textual explanations [11] for each of the process steps. In the following, the group “advanced treatment” refers to the group receiving the advanced assistance system, the group “basic treatment” refers to the group receiving the basic version of the assistance system (without the additional textual explanations). We decided to test two assistant systems with varying amount of text to control for possible differences in the participants usage behavior. The control group received an empty assistance system only stating that the participants should process the emails according to the process specifications. The usage of the assistance systems was voluntary for the participants, but the provided information supported the execution of the experiment’s tasks. We refer to the email client, ticket system, and service catalog as the “experiment’s applications” in the reminder of the paper.

To anonymize the paper, we renamed the assistance system to “ARTIFACTNAME”. In the experiment, the training upfront, and in the survey afterwards, the assistance system had a dedicated name enabling the participants to identify, distinguish, and remember it when reporting their subjective usage. We measured the subjective usage perceptions by two self-developed questions on 7-point Likert scale (intervals in arrow brackets), adapted from the subjective usage measures applied by Davis [4]:

Question use 1: *I used the* ARTIFACTNAME *in the experiment*
<strongly disagree – strongly agree>

Question use 2: *In the experiment, I used the* ARTIFACTNAME
<never – very frequently>

We used Tobii X2-30 eye tracker and the software Tobii Studio for recording the experiment’s sessions. Eye-trackers by Tobii are used in many research domains and Tobii is one of the market leaders of eye-tracking technology world-wide [31]. The eye-tracker recorded the participants’ eye movements with a frequency of 30 Hz. Thus, every 33ms the participants’ eye fixation and position on the screen, for example, was measured and stored with the timestamp by Tobii Studio. In addition, the experiment’s sessions were screen-recorded. The resulting videos as well as the collected data was analyzed subsequently. Following Figure 1 depicts a screenshots of the recorded sessions and the highlighted eye movements of the participants (red dots and lines) based on the collected data by Tobii Studio. The diameter of the red dot indicates the duration of the participants’ eye fixation at the certain point on the screen and the red lines indicate the movement of the participants’ eyes on the screen. Such a visualization of the participants’ eye movements and fixation durations supports researchers to reconstruct the participants’ behavior during the experiment. The screenshot shows a participant with the “advanced treatment” providing the sequence of the ticketing process and the additional explanations for each process step. Please note, to anonymize this paper, the artifact name (top right) and the image of the participant (top left) was blurred.

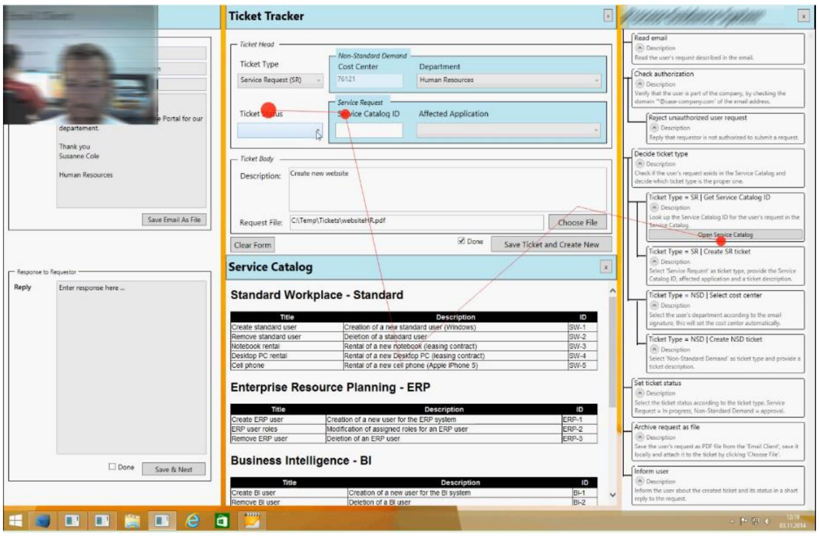

Figure 1: Screenshots of experiment recordings

Within Tobii Studio, we defined four areas of interest (AoI), for the experimental applications (see **Figure 2**). The software analyzes the recordings and provides the participants' fixation counts for each AoI. We used these fixation counts as objective usage measurements grounded by the immediacy assumption and the eye-mind assumption [12].

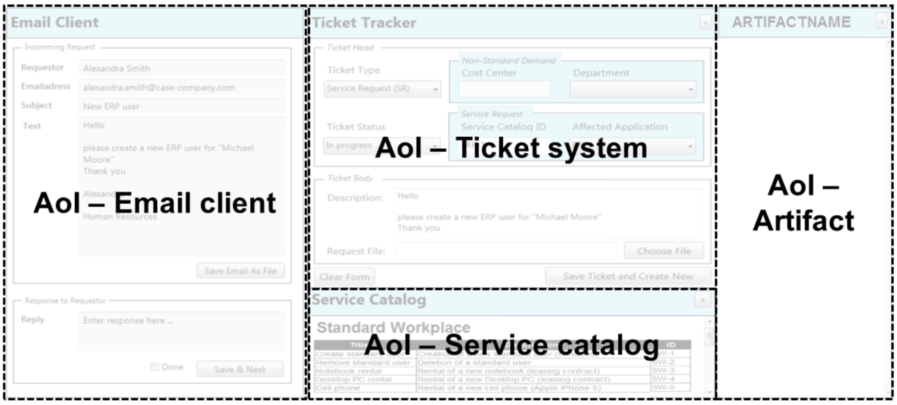

Figure 2: Experimental applications and AoIs

Following Burton-Jones and Straub [15], both usage measurements address the "extent of use" and refer to the second type of measurement richness [15]. Please note we do not discuss the experiment's actual measurements and results here, as we focus on the question whether subjective usage measurements are a reliable approximation of objective usage measurements. However, during the experiment we controlled for the intended effect of the assistance systems and the results indicate a positive effect of both assistance systems compared to the control group. Next, we present the analysis of the participants' subjective and objective usage in the experiment.

5 Analysis

In total, 118 students participated in the experiment, 29 females and 89 males with an average age of 21.81 years (standard deviation = 2.31 years). Table 1 contains the mean values and standard deviations of the total fixations, time, and the survey data for the three experimental groups.

We applied a Shapiro-Wilk test [32] to test if our data follow a normal distribution. All measurements, except for three measures, are not normally distributed, thus referred to non-parametric statistical tests for the data analysis. The mean values of the total fixations indicate a similar usage of the email client, ticket system, and service catalog among the three groups. The artifact's usage varied strongly between the three groups. The highest mean usage, both subjective and objective, has the "advanced treatment" group with a mean count of 2417.97 fixations that equals to 80.59s compared to a mean fixation count of 356.80 fixations or 11.89s of the control group.

Table 1. Mean values and standard deviation of the total fixations, time, and survey data

			Objective measurement					**Subjective measurement**	
Group	n		AoI – Artifact[1]	AoI - Email client[1]	AoI - Ticket system[1]	AoI - Service catalog[1]	Time[2]	Survey - Use 1[3]	Survey - Use 2[3]
Advanced treatment	39	mean	2417.97	6221.92	5634.18	2520.97	715.90	4.41	4.03
		SD	3141.65	1991.18	1905.30	1107.07	288.61	1.85	1.91
Basic treatment	38	mean	1148.60	6357.26	6176.66	2637.37	753.41	3.97	3.50
		SD	1071.33	2598.87	2363.80	1704.95	307.86	2.11	2.01
Control	41	mean	356.80	6899.10	6115.24	2238.24	684.96	2.76	2.54
		SD	297.07	2101.68	2188.53	1312.69	195.76	1.98	1.91

[1] measured in fixations | [2] measured in seconds | [3] measured on a 7-point Likert scale from 1 to 7

As expected, the treatment groups' usage is higher than the control group's usage. In order to gain deeper insights, we calculated the fixations per minutes for the four AoI and visualized them in boxplots (see Figure 3). The results indicate a similar usage of the ticket system and the service catalog among the three groups. Surprisingly, we found a significant difference for the email client usage (measured in fixations per minute) between the "advanced treatment" group and the control group (p-value = 0.016) as well as between the "basic treatment" group and the control group (p-value = 0.012).

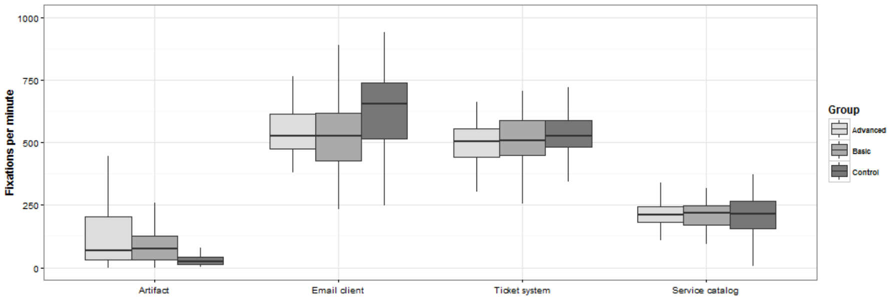

Figure 3. Participants' fixations per minute

The analysis of the artifact's fixations per minute reveals a similar result. For both treatment groups there is a significant difference compared to the control group (both p-values < 0.001). Next, we test for the positive correlation between the objective usage and the subjective usage with the Spearman Rho test [33, 34]. Table 2 contains the Rho values of the correlation tests and the correlations' significance.

Table 2. Correlations between objective and subjective usage (means)

	Complete dataset				**Advanced treatment group**			
	AoI-A	AoI-E	AoI-T	AoI-S	AoI-A	AoI-E	AoI-T	AoI-S
AoI Artifact	1.00 ***	0.46 ***	0.59 ***	0.30 **	1.00 ***	0.66 ***	0.78 ***	0.43 **
Survey - Use 1	0.52 ***	0.24 **	0.28 **	0.19 *	0.68 ***	0.49 **	0.62 ***	0.44 **
Survey - Use 2	0.48 ***	0.24 **	0.26 **	0.21 *	0.60 ***	0.45 **	0.50 **	0.46 **
	Basic treatment group				**Control group**			
	AoI-A	AoI-E	AoI-T	AoI-S	AoI-A	AoI-E	AoI-T	AoI-S
AoI Artifact	1.00 ***	0.57 ***	0.60 ***	0.45 **	1.00 ***	0.40 **	0.60 ***	0.07 n.s.
Survey - Use 1	0.64 ***	0.44 **	0.36 *	0.25 n.s.	0.00 n.s.	0.02 n.s.	0.07 n.s.	-0.08 n.s.
Survey - Use 2	0.59 ***	0.38 *	0.36 *	0.21 n.s.	0.03 n.s.	0.13 n.s.	0.12 n.s.	0.01 n.s.

AoI-A = AoI artifact | AoI-E = AoI email client | AoI-T = AoI ticket system |
AoI-S = AoI service catalog
*** p < 0.001 | ** p < 0.01 | * p < 0.05 | n.s. p > 0.05

The test reveals a correlation between the treatments' objective usage (AoI-Treatment) and the subjective usage measurements (Use1 and Use2) for the complete dataset, the "advanced treatment" group, and the "basic treatment" group. The correlation test for the control group reveals no correlation between the perceived and objective usage measurements. In addition, we tested for a correlation between the subjective usage and the objective usage of the experimental applications. In line with the first correlation test, we found a positive correlation for the complete data set and both treatment groups (except for the "basic treatment" group's service catalog usage).

There is no correlation for the control group. As last step, we tested for a correlation between the objective artifact usage and objective usage of the experimental applications. Here, we found a positive correlation for the complete dataset as well as all three experimental groups (except for the control group's service catalog usage).

6 Discussion

The analysis of the complete data set reveals a positive correlation between the subjective usage perceptions and the objective artifact usage. Moreover, we identified a correlation between the subjective artifact usage and the objective experimental applications usage. This could indicate a type II error, as the participants might not only have assessed their artifact usage, but the overall usage of the experiment's applications. Therefore, we tested for a correlation between the objective artifact's and experimental application's usage. Again, we found a positive correlation. The experiment's setup and task could explain these correlations. The participants used the experimental applications in order to fulfill their experimental tasks and the assistance system supported their process execution. Therefore, there is a correlation between all four objective usage measurements. This fact could also explain the correlation between the subjective artifact's usage and objective usage of the experiments applications. Based on the introspection theory [17], we assume the participants are able to assess their artifact usage, as their cognitive artifact usage is stored in their short-term memory. The objective usage measurements' reliability is based on the immediacy assumption and the eye-mind assumption [12]. Therefore, we can conclude both measurements as reliable and confirm our hypothesis for the entire dataset.

However, the distinct analysis for the three groups reveals contradictious results. Both treatment groups show similar results compared to the complete dataset. However, the control group reveals a different result. There is no correlation between the subjective and objective artifact usage. The control group participants had no information about how to execute the ticketing process provided in their artifact. Thus, we expected that they do not use it (cognitively) in the experiment as their artifact contained only a blank space. Thus, we assume they have the information on "no artifact usage" in their short-term memory. Accordingly, the control group's participants should not report any usage. Nevertheless, the control group reported a low subjective artifact usage (mean value of 2.76 and 2.54).

There are some possible explanations, for example common method biases such as the social desirability bias [14, 35]. The control group participants may had the tendency or need to report some artifact usage in order to "*present themselves in a favorable light*" [14: p. 881], despite the fact they actually did not use it. The participants of the control group could have over-reported their usage because they viewed this behavior as appropriate [35]. Another explanation could be the control group participants were confused and reported their subjective usage of the experimental applications. However, the analysis showed there is no such correlation and therefore, we assume there was no confusion among the participants. Summed up, we can confirm the hypothesis for both treatment groups and reject the hypothesis for the control group.

7 Summary and Future Work

In this paper, we present our research addressing the ongoing debate on subjective vs. objective usage measurements. More specifically, we address experimental settings with cognitive artifact usage, e.g. in the context of DSR projects. In our experiment, we measure the objective usage with eye-tracking technology and the participants' eye fixations. Our analysis supports the assumption that subjective usage measurements are a reliable approximation for objective usage measurements in case of laboratory experiments with an artifact present. When testing a group with an artifact against a group without an artifact the subjective measurements are no reliable approximation for objective measurements according to our analysis. We conclude that researchers could use subjective usage measurements when comparing different artifact designs. In contrast to this, researchers should rely on objective measurements when testing the effect of an artifact compared to a control group with no artifact.

Our research contributes by providing support for the constraint reliability of subjective usage measurements in experimental settings discussed in the body of knowledge. More specifically, our research results indicate the reliability of participants' usage perceptions of an artifact in an experimental setting in case if the artifact is present and can be (cognitively) used by the participants. Moreover, from a methodological point of view, we exemplary show how to measure objective cognitive artifact usage in experimental settings, based on the immediacy assumption and the eye-mind assumption [12]. Eye-tracking technology enables researchers to utilize the participants' eye fixation as a reliable indicator for their cognitive processing and thus, represents an objective usage measurement. Researchers can apply the presented method to assess the participants' (cognitive) artifact usage in their experiments. Especially experiments in the DSR context can benefit from using eye-tracking technology to enrich the usage measurement [15] as the artifact usage is of high importance in DSR experiments [10]. Researchers following the DSR approach can apply the proposed objective usage measurement in their experiments aiming to gain deeper insights into the participants' usage of the artifact and the artifacts' effects.

Although our research follows established methods, there are potential limitations. We conducted the analysis with only one experimental dataset resulting in limited generalizability. There is an ongoing debate on subjective vs objective usage measurements and our findings add to the body of knowledge. Moreover, our study presents a reliable method on how to assess objective measurements for cognitive artifact usage in experimental settings. The analysis presented in this paper uses the aggregate experimental data as our research is still in progress. As next step, we conduct a detailed analysis of the participants' individual usage behavior during the experiment addressing the following activities. **First**, the screen recordings from the eye-tracking software enables us to track the participants' activities and fixations during the experiment, code the participant's usage behavior accordingly, and derive patterns of varying usage behaviors. **Second**, for the presented analysis, we used the total numbers of fixations as the objective measurement. As the used eye-tracking technology measures every 33ms the participant's eye position on the screen, the gathered fixations might also include fixations that occurred, because the participant is scanning the screen (e.g. investigating the applications in the beginning). Based on reading research in psychology, we plan to code fixation clusters with a duration of at least 200 – 300ms, as this time is required to read and comprehend a sentence [25]. This detailed analysis will enable us, for example, to remove the "non-usage" fixations in the experiment's beginning, when the participants scanned the existing applications, in order to get a clearer picture of their actual usage behavior during the experiment. Both research activities will extent the "richness" of our usage measurement following the suggestions by Burton-Jones and Straub [15]. **Third**, we will incorporate the artifact's usage measurements to explain the participants' actual performance in the experiment and test for moderating variables such as gender, ethnicity, and users' knowledge about the experiments ticketing process. Especially the varying level of participants' process knowledge, which increased during the experiment based on the provided assistance, might have an effect on the participants' usage behavior. The detailed analysis of the eye-tracker recordings will enable us to understand the participants changing usage behavior during the experiment and gain further insights on potential learning effects based on the provided assistance.

In addition to the outlined activities addressing our future research, there are further opportunities. We suggest to conduct more studies on the proposed concept of cognitive artifact usage. There is more work required on the theoretical grounding and conceptualization of cognitive artifact usage. In addition, researchers could apply the eye-tracking technology in their experiments to gather objective usage data and provide more findings on the ongoing subjective vs objective usage debate.

References

1. Lucas, H.C., Spitler, V.K.: Technology Use and Performance: A Field Study of Broker Workstations. Decision Science 30, 291–331 (1999)
2. Devaraj, S., Kohli, R.: Performance Impacts of Information Technology. Is Actual Usage the Missing Link? Management Science 49, 273–289 (2003)
3. Eckhardt, A., Maier, C., Hsieh, J.P.-A., Chuk, T., Chan, A.B., Hsiao, J.H., Buettner, R.: Objective measures of IS usage behavior under conditions of experience and pressure using eye fixation data. In: ICIS 2012 Proceedings (2012)
4. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13, 319–340 (1989)
5. Reuver, M. de, Bouwman, H.: Dealing with self-report bias in mobile Internet acceptance and usage studies. Information & Management 52, 287–294 (2015)
6. Barnett, T., Kellermanns, F.W., Pearson, A.W., Pearson, R.A.: Measuring Information System Usage: Replication and Extensions. Journal of Computer Information Systems 47, 76–85 (2006)

7. Straub, D., Limayem, M., Karahanna-Evaristo, E.: Measuring System Usage. Implications for IS Theory Testing. Management Science 41, 1328–1342 (1995)
8. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75–105 (2004)
9. Bhattacherjee, A.: Social Science Research: Principles, Methods, and Practices. Textbooks Collection. Book 3, http://scholarcommons.usf.edu/oa_textbooks/
10. Mettler, T., Eurich, M., Winter, R.: On the Use of Experiments in Design Science Research: A Propositionof an Evaluation Framework. Communications of the Association for Information Systems 34 (2014)
11. Gregor, S., Benbasat, I.: Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. MIS Quarterly 23, 497–530 (1999)
12. Just, M.A., Carpenter, P.A.: A theory of reading: from eye fixations to comprehension. Psychological Review 87, 329–354 (1980)
13. Wu, J., Du, H.: Toward a better understanding of behavioral intention and system usage constructs. Eur J Inf Syst 21, 680–698 (2012)
14. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method biases in behavioral research: a critical review of the literature and recommended remedies. The Journal of applied psychology 88, 879–903 (2003)
15. Burton-Jones, A., Straub, D.W.: Reconceptualizing System Usage. An Approach and Empirical Test. Information Systems Research 17, 228–246 (2006)
16. Nisbett, R.E., Wilson, T.D.: Telling more than we can know. Verbal reports on mental processes. Psychological Review 84, 231–259 (1977)
17. Ericsson, K.A., Simon, H.A.: Verbal reports as data. Psychological Review 87, 215–251 (1980)
18. Dulany, D.E., O'Connell, D.C.: Does partial reinforcement dissociate verbal rules and the behavior they might be presumed to control? Journal of Verbal Learning and Verbal Behavior 2, 361–372 (1963)
19. Bryce, D., Bratzke, D.: Introspective reports of reaction times in dual-tasks reflect experienced difficulty rather than timing of cognitive processes. Consciousness and Cognition 27, 254–267 (2014)
20. Bratzke, D., Bryce, D., Seifried-Dübon, T.: Distorted subjective reports of stimulus onsets under dual-task conditions. Delayed conscious perception or estimation bias? Consciousness and Cognition 30, 36–47 (2014)
21. van Gog, T., Scheiter, K.: Eye tracking as a tool to study and enhance multimedia learning. Learning and Instruction 20, 95–99 (2010)
22. Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S.W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., Tsai, C.-C.: A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educational Research Review 10, 90–115 (2013)
23. Jacob, R.J., Karn, K.S.: Eye Tracking in Human-Computer Interaction and Usability Research. In: Hyönä, J., Radach, R., Deubel, H. (eds.) The mind's eye. Cognitive and applied aspects of eye movement research, pp. 573–605. North-Holland, Amsterdam, Boston (2003)
24. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124, 372–422 (1998)

25. Rayner, K.: Eye movements and attention in reading, scene perception, and visual search. Quarterly journal of experimental psychology 62, 1457–1506 (2009)
26. Traxler, M.J., Long, D.L., Tooley, K.M., Johns, C.L., Zirnstein, M., Jonathan, E.: Individual Differences in Eye-Movements During Reading: Working Memory and Speed-of-Processing Effects. Journal of eye movement research 5 (2012)
27. Marti, S., Bayet, L., Dehaene, S.: Subjective report of eye fixations during serial search. Consciousness and Cognition 33, 1–15 (2015)
28. Johansson, P., Hall, L., Sikstrom, S., Olsson, A.: Failure to detect mismatches between intention and outcome in a simple decision task. Science (New York, N.Y.) 310, 116–119 (2005)
29. Maedche, A., Morana, S., Schacht, S., Werth, D., Krumeich, J.: Advanced User Assistance Systems. Bus Inf Syst Eng (2016)
30. Morana, S., Schacht, S., Scherp, A., Maedche, A.: Designing a Process Guidance System to Support User's Business Process Compliance. In: ICIS 2014 Proceedings (2014)
31. Tobii: About Tobii Pro, http://www.tobiipro.com/about/
32. Royston, J.P.: An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. Applied Statistics 31, 115 (1982)
33. Hollander, M., Chicken, E., Wolfe, D.A.: Nonparametric statistical methods. John Wiley & Sons, Inc, Hoboken, New Jersey (2014)
34. Best, D.J., Roberts, D.E.: Algorithm AS 89. The Upper Tail Probabilities of Spearman's Rho. Applied Statistics 24, 377 (1975)
35. Donaldson, S.I., Grant-Vallone, E.J.: Understanding Self-Report Bias in Organizational Behavior Research. Journal of Business and Psychology 17, 245–260 (2002)