INDIVIDUAL DIFFERENCES IN THE ACTIVE USER PARADOX:

Computer Self-efficacy, Need for Cognition and Affinity with Technology Interaction

RD Timmer, BSC
S1790072
November 3th 2020
Human Factors and Engineering Psychology
University of Twente

Supervisor 1:
Dr. M. Schmettow

Supervisor 2:
Dr. S. Borsci

Abstract

Gaining insight into what individual differences might account for the tendency to fall or resist Active User Paradox (AUP) behavior can provide the opportunity to personalize interventions to help users to become more effective and efficient. Therefore, this study investigates if the traits Need for Cognition, Affinity with Technology Interaction and Computer-Self-Efficacy can account for the tendency to resist AUP behavior. Furthermore, the dynamic nature of CSE was studied, to lay out a basis for further research in AUP behavior and dynamic traits such as CSE. 38 participants completed a computer task which was scored on all AUP related measures (method developed by Keil, Schmettow & Noordzij, 2015) to calculate the scores on the two main factors underlying AUP behavior (cognitive effort and exploratory behavior). Participants filled out multiple questionnaires to measure NFC, ATI and CSE as well. For the explorative analyses, task success and difficulty were measured by two questionnaires. A Principal Component Analyses showed that only one of the two extracted factors corresponded with our theoretical assumptions (exploratory behavior), while the other did not (cognitive effort). Two separate multiple regressions showed that for component two (explorative behavior) none of the traits was a significant predictor and for component one (cognitive effort/knowledge transfer) NFC was the only significant predictor. Furthermore, a morderation analysis showed that Perceived Task Success and Perceived task Difficulty moderate the relation between pre- and post-task CSE. We conclude that methodological issues might account for some of the found null results and that the results should be used for future research in the operationalization of AUP behavior before conclusions can be drawn about the relation between AUP behavior and other constructs.

Table of content

1. Introduction

Interacting with computers has become a part of everyone's life. Although new technologies and computer systems have made us more efficient and effective in daily tasks and life in general, we seem to rarely use systems in the most efficient and effective way. Effectiveness can be defined as the accuracy and completeness with which users accomplish certain goals. This differs from efficiency, in that efficiency concerns the relation between (1) accuracy and completeness with which achieve certain goals *and* (2) the resources expanded in achieving them (FrØkjaer, Hertzum, & Hornbæk, 2000, p. 345). Take, for example, the (relatively new) mobile banking applications, which allow users to manage their bank affairs on their phones. In earlier days, people had to write down all the *exact* transfer information and physically go to the bank before they could transfer money. With the mobile banking function, these people can now transfer money wherever they are with just a few clicks (resulting in less expanded resources), while the application checks if the transfer information corresponds (resulting in more accuracy). Other examples of technologies making us more effective and efficient can be found in updates in office software, as for example MS word. Before the update which included the 'citations' function, users had to type out citations themselves, which could be time consuming when done correctly. When using the citations function, users can now insert a whole citation with one click (resulting is less expanded resources), without the risk of typo's (resulting in more accuracy).

While these examples show that technologies and new functions *enable* us to be more efficient and effective, the question remains: do we *use* these technologies and functions in the most efficient and effective way? Research shows that (most) users tend to fall for the active user paradox (AUP). The paradox of the active user was first described by Caroll and Rosson (1987), and refers to the phenomenon where people use familiar and intuitive procedures for accomplishing a task (i.e. typing out citations), instead of spending time and cognitive effort to explore other procedures that could be more efficient over the long run (i.e. the 'citations' function). Not only keeps this phenomenon users from being more effective and efficient, it is also a thorn in the side for tech companies and developers. Therefore, companies have tried to help their users learn about functionalities (or functions) by, for example, offering users to 'take a tour' through the functions, making them do a trial or introducing a talking paperclip that gives new suggestions. This "help" is mainly based on previous research by Carrol and Rosson (1987) and Fu and Gray, (2004), which focusses mainly on the characteristics of the system (such as the amount of feedback the system provides or the familiarity of the display). Unfortunately, this hasn't led to overcoming the AUP, seeing it is still prevalent (Corbett & Weber, 2016)**.**

A solution – or an useful addition at the least- could be to shift the focus from the system to the user; changing the question from 'what system characteristics might help resisting AUP behavior' to 'What individual differences might help resisting AUP behavior'. Gaining insights into individual differences in AUP behavior could provide us insights in what type of user is more likely to fall for or resist this behavior which could in turn help define a target group; individuals who might need help overcoming the active user paradox, allowing them to become more effective and efficient. Furthermore, the differences between users could form a basis for the creation of effective interventions for the target group (and possibly even subgroups within this group). Given that technology plays such a major role in contemporary life, this could potentially benefit all kinds of individuals in their working as well as their personal life. The current study will therefore use and further explore a newly developed method for measuring individual differences in AUP behavior. The use of this method opens the possibility to answer the main question of our study, namely; can individual differences such as the personality traits (1) need for cognition, (2) affinity with technology interaction and (3) computer self-efficacy account for resisting the AUP in users?

1.2 The active user paradox

Since the introduction of the term by Carroll and Rosson (1987), the active user paradox (AUP) is assumed to arise out of two biases: the production bias and the assimilation bias. The production bias is the desire to obtain results as soon as possible: when people are focused on throughput, they are less likely to invest time in learning new procedures in their tasks, even when these procedures could increase efficiency. The production bias is therefore said to be a *motivational paradox*. The assimilation bias in contrast, is peoples tendency to apply prior knowledge to interpret new situations (Carroll & Rosson, 1987, p1). While this is usually very helpful (imagine entering every new situation as a blank slate), transferring prior knowledge or procedures to a new system might prevent users from "seeing" and using more appropriate or efficient procedures or, even worse, applying wrong knowledge to the new system. Together these biases prevent users from learning new procedures and maximizing their effectiveness and efficiency.

In a more recent attempt to gain insight into the underlaying mechanisms of AUP behavior, Fu and Gray, (2004) were able to identify several characteristics of the procedures that were often chosen by users, but were not the (most) effective and efficient ones. They defined AUP behavior as 'the persistent use of inefficient procedures in interactive tasks by experienced or even expert users when demonstrably more efficient procedures exisist' (2004, p. 901). In their studies, they found that (1) the efficient procedures were often more specialized in contrast to the inefficient procedures (that were often more general). Based on different models of adaptive choice, they argue that the selection for a certain procedure depends on three factors, namely: the frequency of use, the effectiveness of that procedure and the efficiency of that procedure. This feels intuitive: when I use a procedure, experience that it's easy to be accurate and somewhat quick, I will likely use that procedure again. In that way, the use of that procedure is reinforced. In general, procedures that are *more general* (i.e. that are applicable to many programs or problems or tasks) will likely be reinforced more often than procedures that are *more specific,* since the latter only occur in specific programs or problems. One underlying mechanism could therefore be a bias towards general procedures, which tends to be in fact less effective and efficient. A second finding was that the ineffective procedures, that were chosen more often, provided more visual feedback than the lesser chosen effective procedures. This might be quite intuitive as well, if you imagine the following situation. Imagine someone asks you to draw a green circle, place it in the middle of the screen and write your name in it. What is the first thing you do? Do you search for a function that does exactly this ('the-green-circle-name' function)? Probably not. You likely draw a circle and color it green. Once you have a green circle you proceed to the next task, namely placing it in the middle of the screen, and so on. This way, the visual feedback serves as a cue for the beginning of the next part of the task. Fu and Gray (2004) pose that these procedures might require 'less cognitive effort as mental look-ahead can be off-loaded to the external display'(p. 905) and that errors can be noticed quickly. Therefore, they concluded that both biases are in fact cognitive biases and that AUP behavior is a result of underlying cognitive processes of selecting procedures. Following that, the production bias is a result of the advantage of offloading cognitive effort for performance and error control, whereas the assimilation bias is a consequence of cognitive availability of- and experience with general procedures.

While Fu and Gray (2004) offered a framework to research the occurrence of the AUP and new insights on the characteristics of the procedures related to the AUP, some questions remain unanswered. For example, what makes that some participants used inefficient procedures while others used efficient procedures? Furthermore, Fu and Gray (2004) as well as Carroll and Rosson (1987) only differentiate between users based on their experience (experts, trained participants and novices). But what differences between users matter in AUP related behavior? This leads us to the question what individual differences might account for the tendency to resist the active user paradox.

1.3 Individual differences in resisting the Active User Paradox

The first to explore individual differences in relation to the AUP were Keil, Schmettow, & Noordzij (2015). They studied individual differences in the tendency to fall for AUP behavior. The AUP was defined as 'the unwillingness to invest time and cognitive effort for the exploration of procedures' (p. 9). Resistance to the AUP on the other hand, was defined as the willingness to do so (Keil et al., 2015). Furthermore, they adapted the method of Fu and Gray (2004) so that it would be suitable to detect individual differences in AUP related behavior, instead of focusing on properties of procedures. This led to the creation of a method in which participants were asked to perform basic computer tasks in a graphic program (three tasks) and in MS word (2 tasks), which were recorded and then scored on AUP behavior measures. In addition, they targeted participants' with basic computer experience, in contrast to experts of computer programs and trained participants.

The definition indicated two factors that represented AUP resistance: exploratory behavior and invested cognitive effort. Therefore, high scores on these components were expected to indicate high AUP resistance. Invested cognitive effort was assessed by identifying and scoring the procedures used by participants, according to the GOMS model (John & Kieras, in Keil et al., 2015). Scoring was done by assessing five characteristics of the procedures that were expected to increase the required cognitive effort, based on the previous research by Fu &andGray (2004). These five measures were specificity, difficulty, complexity, delayed feedback and parameter demands. Exploratory behavior was assessed with four different measures, namely; method diversity, read-handout duration, undo amount and set-parameter amount. As expected by the researchers, the findings showed individual differences in AUP resistance behavior between participants. Not *all* developed tasks seemed suitable for the detection of individual differences and not *all* behavior measures showed sufficient variance (see method section for our adjustments), but all tasks and measures that did were included in the analysis. The results were mixed: on the one hand, they succeeded in developing a measure that allowed to measure individual differences in AUP (related) behaviors, on the other hand, there were some finings that were unexpected and ask for a follow-up study to optimize the method and design, to see if the same results will be found. One important finding was that a factor analysis had shown that the AUP behavior measures landed on two components. One was defined as exploratory behavior, but the other did not seem to theoretically represent cognitive effort. Furthermore, while efficiency and effectiveness measures were (as expected) related to several AUP behavior measures, the absence of some significant relations with other measures (such as the cognitive effort component and the exploratory behavior component), raised the question if the task was repetitive enough for individuals who show AUP resistance behavior to gain efficiency and effectiveness compared to the others. Finally, *Need for Cognition* (NFC), was shown to be (weekly) associated with higher AUP resistance, while *Geekism* was not.

The current research aims at further exploring what traits are involved and from there on better understand the cognitive processes that cause AUP behavior. By adapting the designed method for the detection of individual differences, we will assess if we can replicate the found role for NFC, and explore other characteristics, namely affinity with technology interaction and self-efficacy, that could potentially prevent individuals from falling for the Active User Paradox.

1.4 Affinity for Technology Interaction & Need for Cognition

Keil et al. (2015) measured two personality characteristics to see whether they contribute in resisting the AUP. Need for Cognition (NFC), the tendency to engage in and enjoy complex thought (Cacioppo & Petty, 1982, as cited in Sadowski, 1993) was believed to be related to AUP resistance because individuals with a high NFC tend to show more cognitive flexibility in their choice of learning strategies, are usually highly motivated for challenging tasks and have excellent control over their attention resources (Keil et al., 2015). Geekism, defined as a predisposition that is associated with great affinity for exploring and tinkering with technological devices (Schmettow, Noordzij & Mundt, 2013), was hypothesized to be related to AUP resistance, because individuals with a high geekism score tend to focus on acquiring computer skills instead of throughput. Although both traits showed some associations with AUP resistance measures, only NFC reached statistical significance. This was unexpected, since a relation was assumed, based on theory.

Recently, Franke, Attig and Wessel (2018) created and validated a scale, which is rooted in NFC and closely linked to Geekism: the affinity for technology interaction scale (ATI). Users ATI is defined as the extent to which users tend to actively approach interaction with technical systems or, rather, tend to avoid intensive interaction with new systems (Franke et al., 2018, P. 1). The objective for creating the ATI scale, was to create a scale that focuses on de *interaction facet* of affinity for technology (instead of attitudes, for example) and can differentiate over the whole range of the trait, where geekism can be seen as a more extreme part of the full dimension. For these reasons, AIT might be a better predictor for AUP resistance behavior dan geekism. Furthermore, it could be argued that ATI is an even better predictor than NFC was found to be because, while they both emphasize the enjoyment of complex thinking, ATI focusses only and specifically on technical systems instead of more general situations in which one might enjoy complex systems or thought. For example, an item of the NFC scale is 'abstract thinking is something I enjoy', whereas the ATI scale states 'I enjoy spending time getting familiar with a new technical system'. This study will therefore explore if ATI is indeed more strongly related to AUP behavior than NFC is, and if it is a better predictor.

1.5 Self-efficacy

Since the work by Bandura (1977), self-efficacy holds an important role in psychology and has received attention in different domains such as learning, school psychology, sports psychology and psychopathology. In his Social Cognitive Theory (Bandura, 1991), Bandura describes motivation and behavior as outcomes of interacting determinants. These determinants – personal factors, environmental factors, and behavior- influence each other bidirectionally. Because of the changing nature of these factors, people use self-regulatory processes to monitor these factors and “update” their strategies, affect, cognitions and behavior accordingly (Schunk et al., 1999). Amongst these processes, self-efficacy is the most influential one. Self-efficacy is defined as ‘an individual’s perception of both his or her ability and capability to execute as well to achieve successful behavioral outcomes (Marks & Allegante, 2005). Individuals with high self-efficacy beliefs, have shown, amongst other things, to be more persistent when faced with difficulties and they sustain their effort after setbacks. They quickly recover their sense of self-efficacy after setbacks or failure. Individuals with low-self-efficacy in turn, are prone to withdraw from difficult tasks. An individual’s self-efficacy beliefs are believed to derive from one or more of the following sources; performance accomplishments, vicarious experiences, verbal persuasion and physiological states. Performance accomplishments are the primary source of self-efficacy beliefs. Successful accomplishments increase self-efficacy beliefs, while unsuccessful experiences weaken or undermine self-efficacy beliefs, but only when there are no or little previous successful experiences to fall back on.

Over the last years, self-efficacy had gained interest from HCI researchers, which led to a domain specific self-efficacy construct: computer self-efficacy (CSE). CSE is made up by two levels: (1) a general CSE (‘I can successfully use a computer’) and (2) application-specific self-efficacy (‘I can successfully use application X’) (Yi & Hwang, 2003). Just as in Bandura’s self-efficacy theory, computer self-efficacy is believed to influence perseverance when faced with difficulties and challenges. We wonder if individuals with high computer self-efficacy beliefs are more likely to resist the AUP, because their likeliness to persevere when faced with difficulties might facilitate explorative behavior. In contrast, individuals who give up or stop their search when a strategy or function does not seem to work (low CSE), will probably never explore (which is often characterized by trial and error). Furthermore, individuals who persevere (high CSE) probably invest more cognitive effort in the task than individuals with who don’t persevere (low CSE), based on our assumption that persevering when faced with difficulties will require more cognitive effort than quitting the task or exploration. Individuals with high self-efficacy beliefs might therefore be expected to resist to fall for the AUP more often than individuals with low self-efficacy beliefs.

One question that rises is how to measure a dynamic trait like CSE. Because the mere execution of the repetitive tasks in the AUP method can be seen as multiple performance accomplishments, could it be possible that CSE beliefs change during the experiment? Several studies examined the development of CSE in computer based environments. They too, emphasize the importance of past experiences. However, it is not the quantity of the prior experience that is a direct determinant of CSE, but rather the quality of those experiences. In other words, it is not about how many times a program or computer has been used before, but rather how it’s used (successful or onsuccesful). Chang & Chiou (2016) studied the dynamic nature of CSE, by using a circuit design. Students CSE levels where measured before and after a 20 minute computer task. Even after this short amount of time, the results showed increases and decreases in CSE, depending on pre-CSE self-efficacy levels and valence of the task experience. Therefore, they conclude that both prior self-efficacy levels, as well as valence of task experience affect their changes in self-efficacy.

Since the used method in this study includes performing tasks in a relatively unknown graphic program, participants will likely not have a lot of previous experiences (performance accomplishments) to base their self-efficacy beliefs on. Therefore, we expect the first encounter with the application will have a major influence: it will decrease self-efficacy beliefs when using the application is too difficult or hard (unsuccessful performance accomplishment) and increase when using the application leads to the desired results (successful performance accomplishment). This influence will likely be much smaller for participants who already have high CSE beliefs prior to the task, since they are characterized by quick recovery from setbacks. We assume that this will be reflected in a lesser effect on post-CSE (compared to individuals with low pre-CSE), even when they have a negative task experience. Therefore, we propose that self-efficacy beliefs can increase or decrease during computer tasks, especially when the tasks are repetitive, when individuals have no prior experience with the specific computer program and when they have low CSE beliefs at the beginning of the task. Therefore, self-efficacy should be measured pre- and post-task, as it is likely that self-efficacy beliefs differ between those time-points. Figure 1 and Table 1 show a visual representation of the model and expectations.

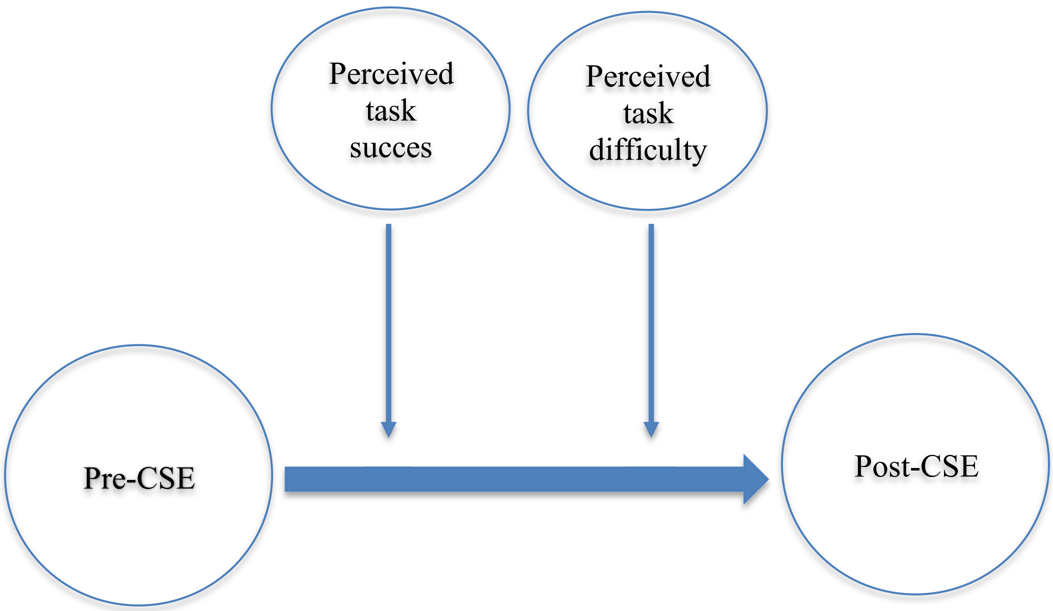


Figure 1.

Visual representation of the proposed theory about the dynamic nature of CSE and the influence of perceived task success and perceived task difficulty

Table 1

A visual representation of the described expectations concerning the change in CSE levels and the effects of perceived task success and perceived task difficulty.

Pre-CSE	Perceived task succes	Perceived task difficulty	Post CSE
Low	Low	High	Lowest
	Low	Low	Low
	High	Low	Highest
	High	High	High
High			Small or no effect

Note. The categorization low-high is not defined by their absolute difference, but rather the relative difference, since we have no information about what 'absolute' levels of low or high CSE are.

1.6 Research questions

In summary, the aim of this research is threefold. Firstly, we want to replicate the study by Keil et al., (2015) to see whether the developed measure is indeed able to detect individual differences in AUP resistance behavior. In doing this, we will include some proposed improvements to the method (as described in the method section). If the measure is able to detect individual differences, this should be reflected by sufficient systematic variance in AUP resistance measures and a factor analyses on these measures should reveal two underlying factors; invested cognitive effort and explorative behavior. Furthermore, as we expect individuals with high AUP resistance behavior to be more efficient and effective in the end, a positive relation should be found between those constructs. Because our hypothesis is that with our adjustments to the developed measure, we will find the measure to be suitable for detection of AUP behavior, our second aim is to explore if NCF, ATI and CSE can account for the tendency to resist AUP behavior. Based on previous research and literature, we expect all three personality traits to be predictors of this behavior and that ATI will be the strongest predictor out of the three. Lastly, we add an explorative component to our research by investigating the dynamic nature of CSE through a moderation, which might form a basis for further theorization about the relationship between AUP behavior and CSE. We expect participants' CSE levels to decrease or increase between pre- and posttest measurements, depending on their perceived task difficulty and perceives task success. Low perceived task difficulty combined high task success will lead to an increase of CSE levels and high perceived difficulty combined with low perceived task success will lead to a decrease of CSE levels.

2. Method

2.1 Participants

Participants were recruited in and around the University of Amsterdam, by an email to staff and students from the second year of the Bachelor Psychology and a facebookpost. Exclusion criteria were taken over from Keil et al., (2015): color blindness and less than basic computer experience. Only participants between the age of 18-35 where selected, to ensure a basic knowledge of computer usage. No incentive was offered for participation. The abovementioned procedure resulted in a sample of 38 participants.

2.2 Apparatus

The tasks were based on the study of Keil et al. (2015), where five tasks were developed to measure AUP behavior. In the first three tasks, participants were asked to manipulate images in the graphic program *GIMP* (version 2.8.14). To illustrate, one of the task instructions was: ‘Change the color of all green objects and turn them blue, without damaging the other objects. The fourth and fifth task where text-formatting tasks, using *Microsoft word*. Each task could be carried out with different procedures, using different tools. Participants were provided with a tool-guide for the *GIMP* task, but free to choose if they would use this and what procedure they carried out. Every task consisted of five similar repetitions (except task 5), creating a situation where it might be more efficient to invest time in exploration of procedures than to hold on to intuitive procedures.

The GIMP files for the tasks where taken over from (Keil et al., 2015), but we made two adjustments:

1) Task two (graphic task) seemed unable to show individual differences. Therefore, a new and more challenging version of task two was created. In the new version, participants were instructed to change the color of the figure into the background color and to change the background color into the color of the figure, without changing the position of the figure. For a visualization, see figure 2. Because adjustments did not ask for modifications in the *GIMP* settings used in the original experiment (for example, what tools are available), we used the exact same settings (see Keil et al (2015) and Appendix 7.3).

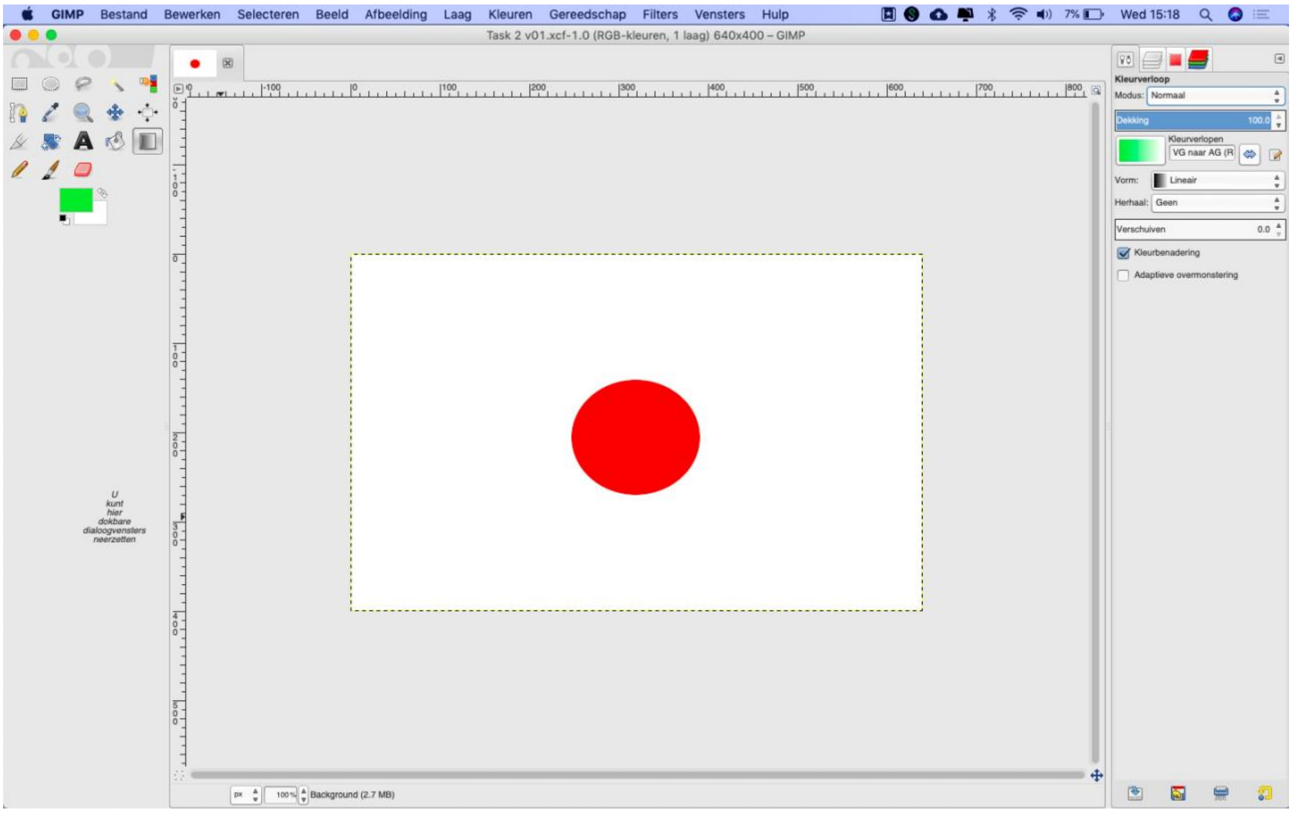


Figure 2. Example of one of the versions of the newly constructed task 2.

2) The text-formatting tasks showed low method diversity and where therefore unable to measure individual differences. Although this is worth further exploration, we have chosen to delete the text-formatting tasks, because the detection of individual differences is the primary interest of our study.

2.3 Measures

2.3.1 Need for cognition

Need for cognition (NFC) was measured using the Need for Cognition Scale (Cacioppo, Petty&Kao, 1984; Keil, Schmettow, & Noordzij, 2015). The questionnaire consists of 18 items, which are scored on a 6 point Likert scale (0= strongly disagree, 6= strongly agree). An example of a positively worded item is ‘I really enjoy a task that involves coming up with new solutions to problems’. An example of a negative worded item is ‘Learning new ways to think doesn’t excite me very much’. The total score is calculated by averaging all item scores after reversing the scores of the negative items (items 3, 4, 5,7, 8, 9, 12, 16 and 17). A high score indicates a high NFC. The NFC scale has a high internal consistency (theta of +.90). The full questionnaire can be found in Appendix 7.7.

2.3.2 Affinity for technology

Affinity for technology (ATI) was measured using the ATI Scale (Franke, Attig & Wessel, 2018). The questionnaire consists of 9 items. Response scale options ranged from 'completely disagree' (=1) to 'completely agree' (=6) on a six- point scale. An example of an item is: 'I like testing the functions of new technical systems'. The total ATI score is calculated by averaging all item scores after reversing the scores of the three negatively worded items (items 3, 6, 8, reversed to 6=1, 5=2, 4=3, 3=4, 2=5, 1=6). A high score indicates a strong affinity for technology. The internal consistency of this scale is good to excellent (Cronbach's alpha ranging from .83-.94.) and it has a good convergent- and discriminant validity. The full questionnaire can be found in Appendix 7.6.

2.3.3 Demographic variables

Demographic variables were measured using an extra questionnaire. Gathered variables included gender, age, study or profession, and frequency of computer use (daily/a few times per week/ monthly). Furthermore, participants had to indicate their knowledge of graphical programs on a 10-point scale (0= no knowledge, 10 = high level of knowledge).

2.3.4 Prior experience

Prior experience was measured by two 5-item questionnaires. The first questionnaire asked participants how often they have used the graphic programs *Microsoft paint, Adobe photoshop, MacPaint, GIMP* and/or *Paintbrush (*never =0 /seldom = 1/monthly or more =2), indicating the frequency of the experience (no experience, low experience, high experience). Because we are interested in the overall experience with graphic programs, the sum of responses over all 5 items was calculated to derive one experience score per participant.

The second questionnaire asked participants how much they enjoy using these programs (not at all =1/ not very much =2/ a little =3/ very much=4) indicating the valence of the experience (strongly negative, slightly negative, slightly positive, strongly positive). An average score was calculated to get one valence score per participant.

2.3.5 Computer Self-efficacy

One of the most common used CSE questionnaires in the HCI domain is constructed by Murphy, Coover, and Owen (1989). They created 32 items with a 5-point likert-scale (1= very little confidence, 5= quite a lot of confidence) with three underlying factors: beginning level computer skills, advanced level computer skills and mainframe computer skills. Although this questionnaire is widely used, it is commonly modified to fit the rapidly changing HCI domain. First, we replaced the

5-point likert-scale by a 7 point likert-scale because people usually avoid the extremes, which makes that small scales are not able to detect individual differences that well. Secondly, we changed the instructions so that participants were instructed to rate their confidence level about executing the behavior *now,* and not future capabilities (which was based on Bandura's guide for constructing self-efficacy scales, 2005). Furthermore, we excluded all items on de mainframe computer skills factor, and items 3, 4, 7, 8, 17,21 and 27. Several items where reworded to fit the usage of computers in 2018. An example of an item is '*I feel confident troubleshooting computer problems'.* The complete questionnaire can be found in the Appendix 7.8. The constructed questionnaire consists of 23 items, with a response scale ranging from 1(cannot do at all) to 7 (highly certain can do). The total CSE score is the sum of the rescored item scores dived by the number of items, resulting in a total score between 1-7. A high score indicates high self-efficacy beliefs. Assuming we only 'updated' the items, and not change the content, we expect a high internal consistency which will approach the internal consistency scores of the original questionnaire (alpha's between .92-.97). To make sure participants answered the post CSE questionnaire based on their current beliefs, and not on their recall of the answers they gave on the pre CSE questionnaire, the CSE-questionnaire-b presents the items in randomized order. Pre-task CSE questionnaire is labeled 'CSE-a', whereas post-task CSE questionnaire is labeled 'CSE-b'.

2.3.6 Perceived task success

Hornæk (2006) provides an overview of measures used in HCI research. Several studies measure the subjective construct 'perception of outcomes' or related constructs in different ways. To our knowledge, there is not one questionnaire the field that measures perceived task success that is widely used. Therefore, we choose to create our own scale to measure perceived task success. The questionnaire was created with three factors (success, efficiency and individual approach), which led to 11 items. An example of an item is 'I completed the tasks relatively quickly'. The items were scored on a 5-point Likert scale which were added up to come to the total perceived task success score, ranging from 0 (=task perceived as not successful) to 55 (task perceived as highly successful). The perceived task success questionnaire was presented with the perceived difficulty questionnaire.

2.3.7 Perceived task difficulty

Perceived task difficulty was measured with a 4 item questionnaire with a 5 point Likert-scale response scale (0= not difficult, 5= very difficult). An example of an item is 'The task was difficult to complete successfully'. Before adding up the item scores to create a total perceived task difficulty score, items 1 and 2 had to be rescored (5=1, 4=2, 3=3, 2=4, 1=5) because these items were negatively worded. The total score then ranged from 0 to 20, meaning 'not difficult' and 'very difficult' respectively. The perceived task success questionnaire and the perceived difficulty questionnaire are presented as a combined questionnaire, referred to as 'task-reflection questionnaire', as presented in Appendix 7.8.

2.4 Procedure

After taking place behind the laptop, participants were informed about the aim of study. In order to minimize the effects of biases, the description was provided vaguely: 'the aim of the study is to acquire insights in a new method for data-analyses in usability testing.' They were also informed that their computer use would be recorded. All participants who wanted to proceed were asked to read and sign the informed consent (Appendix 7.4). Before the experimenter left the room, she explained the procedure to the participants (including how to switch between the tool guide and *GIMP,* how to save files) and handed them the hand out (including the demographic and previous experience questionnaire and instruction hand-out). Next, screen recording with *Quicktime screen recording* was started by the experimenter. All necessary files for the tasks could be found on the desktop. The steps of the procedure where described in the instruction handout (Appendix 7.2). This way, the participants could work through the steps independently and at their own pace. The order of the steps was; (1) Demographic- and previous experience questions, (2) ATI scale, (3) NFC scale, (4) CSE-a, (5) task 1 instructions, (6) task 2 instructions, (7) task 3 instructions, (8) CSE-b, (9) task reflection questionnaire. After filling out the last questionnaire, participants notified the experimenter. From here, participants were debriefed about the aim of the study.

2.5 Scoring

Methods, cognitive effort, exploratory behavior, efficiency and effectiveness were scored in the same way done by Keil, Schmettow, & Noordzij (2015). An overview can be found in their article and in table 2. While they concluded that not all measures seemed suitable for measuring individual differences in resisting the active user paradox (such as delayed feedback and cognitive effort), we will not modify the measures and their scoring approach for this study, because we will try to replicate their design as much as possible.

Table 2

Measures and their scoring approach by Keil, Schmettow, & Noordzij (2015)

Measure	Scoring approach
Method characteristics (specify, difficulty, complexity, parameter demands and delayed feedback)	Methods are scored into the five method characteristics according to (low = 1, medium = 2, high = 3) and the sum of each of the 5 method characteristics in a subtask is divided through the amount of used methods in this subtask
Cognitive effort	Sum of all the method characteristic scores with sufficient variance divided through the number of characteristic scores with sufficient variance
Read-handout duration	Duration of the GIMP tool guide being in the foreground of the computer screen during a subtask
Undo amount	Sum of uses of the undo function during a subtask
Set parameter amount	Amount of times parameters were modified before using a method
Method diversity	Amount of different methods used per subtask
Efficiency	Time on task, amount of methods used per subtask
Complete faults	Sum of all items that were not modified according to the task description plus all items that were completely modified but should not be modified
Half faults	Sum of all items that were only partially modified according to the task description plus all items that were modified partially but should not be modified
Items to be modified	The amount of items that had to be modified plus 0.5 for each half fault and plus 1 for each complete fault.
Effectiveness	1 - (complete faults/Items to be modified)-(half faults/items to be modified/2)

4 Results

4.1 Participants

All 38 participants were between the age of 20 – 34 (*M*= 27.13, *SD* = 3.55). The sample included 18 men and 20 females. Professions of participants were diverse, ranging from students and teachers to a vlog editor and a PR manager. All but one participant indicated that they use a computer daily, where the one participant indicated 'seldom'. Since this participant did use a computer daily when she studied (a 4 years ago), we assumed she had basic computer experience and included her in the study.

4.2 Variance AUP measures

Means and standard deviations were calculated for all AUP behavior measures, to see which showed appropriate variance. As can be seen in table 3, for the cognitive effort measures, four of the five measures show some level of variance. The exception is delayed feedback, which showed no variance (*M*=1, *SD*= .00) in any of the tasks. The scores of the measures had a range from 1-3, therefore specifity, difficulty, complexity and parameter demands can be said to show moderate variance in task 1 and 3 and only little variance in task 2 (*SD*= .15, .16, .29, .18). For the abovementioned reasons, we have chosen to exclude the delayed feedback measure and continue the analysis with task 1 and 3.

Table 3

Means and standard deviations for all AUP behavior measures.

		Task 1	Task 2	Task 3
Cognitive effort	Specifity	2.00 (*.38*)	2.09 (*.15*)	2.09 (*.26*)
	Difficulty	1.27 (*.28*)	1.18 (*.16*)	1.30 (*.30*)
	Complexity	2.67 (*.36*)	2.80 (*.29*)	2.69 (*.34*)
	Delayed feedback	1.00 (*.000*)	1.00 (*.000*)	1.00 (*.000*)
	Parameter Demands	1.68 (*.29*)	1.80 (*.18*)	1.77 (*.24*)
Exploratory behavior	Method Diversity	4.18(*1.52*)	4.34(*2.08*)	4.43 (*1.66*)
	Read handout	24.97(*32.48*)	32.97 (*70.05*)	8.51 (*21.32*)
	Set Parameter amount	2.79 (*3.34*)	1.82 (*3.95*)	4.95 (*6.78*)
	Undo amount	1.58 (*2.48*)	9.66 (*12.81*)	3.57 (*4.78*)

4.2 Factor analyses AUP measures

A principal axis factor analysis was conducted on the 8 AUP resistance behavior measures with oblique rotation. The Kaiser-Meyer-Olkin measure verified sampling adequacy for the analysis, KMO = .715. Although the factor analysis initially showed three factors with an eigenvalue above Kaisers criterion of 1, together explaining 72.54% of the variance, we proceeded with only the two major factors, because of our theoretical predictions. The first two factors explain 59.42% of the variance. Table 4 shows the factor loadings after rotation.

Table 4

Factor loadings for the AUP behavior measures: cognitive effort (complexity, difficulty, specifity, parameter demands) and exploratory behavior (method diversity, RH duration, set parameter amount, undo amount)

	Component 1	**Component 2**
Complexity	-.872	
Specifity	.868	
Parameter Demands	-.840	
Method diversity	.722	
Set parameter amount		.760
Difficulty	.394	.653
RH duration		.624
Undo Amount		.546

Based on the theory the study of Keil et al. (2015), we expected two factors to be extracted: cognitive effort and exploratory behavior. We largely replicated their findings. The measure Complexity seems to be strongly negatively related to the first factor, meaning that individuals scoring high on this factor, tend to use methods with few tool options or no available parameters. Not surprisingly, the measure Parameter demands is strongly negatively correlated as well, indicating that that these individuals used methods that can easily be used without changing the parameters. Furthermore, the measures Specifity, Method Diversity and Difficulty showed moderate to highly related to the first factor. In other words: the first factor seems to be related to using methods that suitable for specific tasks and objects, using a greater variety of methods and more difficult methods, without using methods that require changes in settings or parameter demands for effective use. While it was originally expected that these measures would measure the factor ‘cognitive effort’, this does not seem in line with the factor loadings and the interpretation of the measures. Keil et al (2015) suggested this first factor could be interpret as ‘knowledge transfer’ from previous experience. Mainly because this factor is related to using a lot different and specific methods without being related to the use of undo command or read handout duration. To find substantiation for this assumption, we (exploratively) looked at the correlation between participants' scores on the first factor and (1) their frequency of experience with other graphic programs and (2) their own indication of their knowledge of graphical programs (between 1-10). While we did find a significant correlation between the two indicators of knowledge (r=.795, p < .001), no significant correlation was found between the first factor and frequency of experience (r=.131, p=.439) or the first factor the graphical knowledge scores (r=.278, p=.091). Based on these results, the first factor does not seem related to knowledge transfer.

The second factor does seem to be according to expectations: the measures undo amount, read handout duration and set parameter amount seem to be positively related to the second factor, which can therefore be interpret as explorative behavior. Whilst the measure difficulty wasn't expected to be related to the explorative behavior, a moderate relation (r=.737) was found. This is not in concurrence with the results of Keil et al (2015), but might not be that surprising: individuals who are actively exploring a program are likely to find (or stumble upon) more difficult methods.

4.3 Efficiency and effectiveness

To see if the individuals who show more AUP resistance behavior (the two extracted factors) are more effective and efficient in the tasks, a Pearson correlation on the two extracted factors, measure of effectiveness and measures of efficiency (time on task and method amount) showed that, in contrast to our expectations, only three significant relationships exist between these variables. The first is a relation between the efficiency measures time on task and method amount (r= .64, p<.001), which was expected because both are indicators of the same construct. In this case, a high score on method amount and time on task implicate low efficiency. The second significant relation was found between the effectiveness measure and time on task (r=-.352, p=-.03). The negative relation implicates that individuals who were more effective (made less mistakes) spend more time on the task. The third found relation was a negative relation between 'method amount' and the second extracted factor (explorative behavior), r=-.33, p=0.04. This indicates that participants who showed more explorative behavior (who read the handout, had a high undo amount, used more difficult methods), used less methods to complete their tasks. While is it appealing to conclude that people who show more explorative behavior are indeed more efficient, the relations we *did not find* might indicate that this conclusion is too simplistic and maybe even incorrect. Given that participants with high scores on explorative behavior used less methods but did not significantly spend less time on the tasks than others did and were not significantly more effective, we can't yet conclude they were more effective than others.

4.4 NFC, ATI & (pre) CSE

A multiple regression was conducted to inspect the effects of NFC, ATI and CSE (pre) on the first found component. The correlation matrix shows that pre-CSE correlates significantly to both NFC (r =.504, p= .001) and ATI scores (r= .631, p < .001). Furthermore, NFC (r= .422, p= .004) as well as pre-CSE (r=.296, p= .036) seem to correlate significantly to the first component. Since these are interpreted as moderate correlations, we assume there is no multicollinearity (all R's <.9). VIF scores substantiate this assumption, with VIF scores ranging from 1 – 2.2. Because other assumptions don't seem to be violated, we interpret the regression analysis without bootstrapping. The results show that NFC seems to be the only predictor of the first component, accounting for 19,1% of the variance in our sample. Adding ATI and pre-CSE to the model only adds 4% and .001% respectively to the amount of explained variance of our sample. The Adjusted R Square (.018, .121) shows that if we would generalize to the population, this contribution would even be less. Looking at the model parameters, we can only conclude that NFC is making a significant contribution to the model (with a slope of .200 - 1.828 in 95% of the times). The results can be found in table 5.

Table 5

Linear model of predictors (ATI, NFC, pre-CSE) of component 1 with confidence intervals.

		b	**SE B**	**B(s)**	**lower**	**upper**
1	Constant	-.727	.582		-1.908	.454
	ATI	.201	.115	.211	-.114	.515
2	Constant	-4.353	1.537		-7.472	-1.233
	ATI	.113	.149	.119	-.189	.415
	NFC	1.024	.406	.394	.200	1.828
3	Constant	-4.337	1.560		-7.507	-1.167
	ATI	.089	.191	.093	-.298	1.927
	NFC	.978	.467	.376	.029	1.927
	Pre-CSE	.048	.229	.047	-.417	.513

A second multiple regression was conducted to inspect the effects of NFC, ATI and CSE (pre) on the second found component. The correlation matrix shows that none of the predictors correlates significantly to the second component. Furthermore, none of the models seem to include a significant predictor of the second component, with all models explaining less than 1% of the variance. The model parameters can be found in table x.

Table 6

Linear model of predictors (ATI, NFC, pre-CSE) of component 2 with confidence intervals.

		B	SE B	B(s)	Lower	upper
1	Constant	.201	.523		-.861	1.263
	ATI	-.055	.139	-.066	-.338	.227
2	Constant	-.496	1.496		-3.533	2.542
	ATI	-.072	.145	-.086	-.366	.222
	NFC	.197	.395	.086	-.606	.999
3	Constant	-.520	1.517		-3.604	2.563
	ATI	-.033	.185	-.039	-.409	.344
	NFC	.272	.454	.119	-.652	1.195
	Pre-CSE	-.078	.223	-.087	-.530	.375

4.5 Exploratory analyses

A moderation analyses on the pre-CSE- and post-CSE scores with perceived task success and perceived task difficulty as moderators, showed that both moderators showed a significant interaction effect, b= -0.25, 95% CI[-.46, -.039], t=-2.42, p=.022 and b= -.16, 95% CI[-.31, -.02], p=.03 respectively. The results of the full model can be found in table 7. Figure 3 shows that when pre-CSE scores are high, there is no effect from perceived task success on post-CSE scores and only little effect from perceived task difficulty (where higher task difficulty predicts lower post-CSE scores). Interestingly, the effects of the moderators seem to increase when pre-CSE scores are low. For example, when pre-CSE is low, perceived task success is low and perceived task difficulty is low, the negative effect on the post-CSE scores is the strongest, compared to the other levels of the moderators. This is not in line with our expectations, because we expected to see the strongest negative effect when perceived task success was low and perceived task difficulty was high (which corresponds to a situation in which the task felt unsuccessful and very difficult to perform). In contrast, the figure shows that when pre-CSE is (relatively) low and perceived task success is low, post-CSE scores are generally lower when perceived task difficulty is low, compared to when perceived task difficulty is high. So, it seems that for these individuals, a situation in which they feel unsuccessful while the task is seen as not that difficult, has the greatest negative impact on their CSE beliefs. On the other hand, in case of low pre-CSE and high perceived task success, high perceived task difficulty seems to have a greater positive effect on post-CSE than low perceived task difficulty. This too was not fully according to our expectations, because we expected to see the greatest positive effect in this situation when task difficulty was low.

Table 7

Linear model of predictors (pre-CSE, perceived task success, perceived task difficulty and their interaction effects) on post-CSE.

	b	***SE B***	***t***	***p***
Constant	5.27	.06	83.00	< .001
	[5.14, 5.40]			
pre-CSE	.83	.08	10.85	< .001
	[.68, .99]			
Perceived task success	.19	.11	1.72	.096
	[-.04, .41]			
Pre-CSE x Perceived task success	-.25	.10	-2.42	.026
	[-.46, -.04]			
Perceived task difficulty	-.05	.07	-.71	.485
	[-.20, .09]			
Pre-CSE x perceived task difficulty	-.16	.07	-2.27	.030
	[-.31, -.02]			

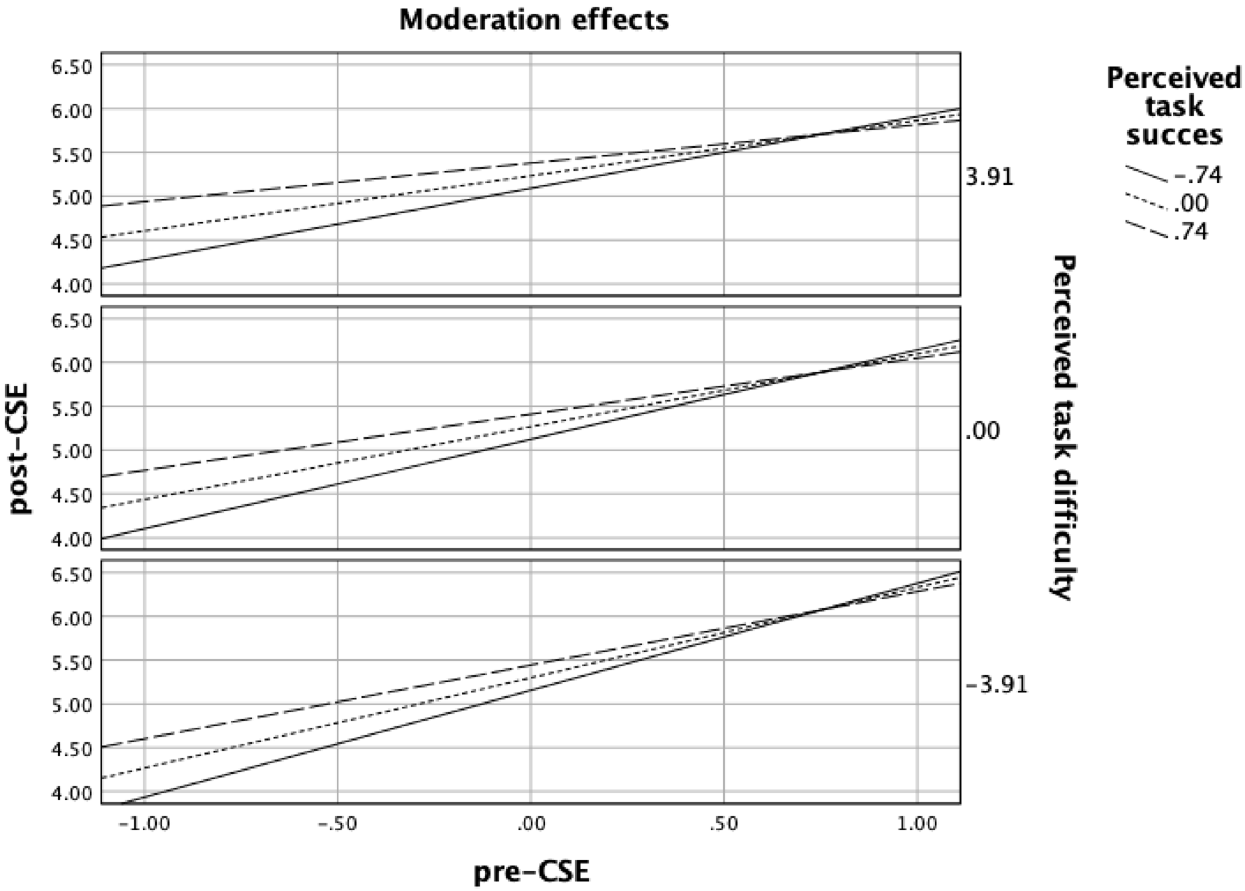


Figure 3. Plot of the predictors (pre-CSE, perceived task success, perceived task difficulty and their interaction effects) on post-CSE.

5. Discussion

5.1 Main conclusions

We have studied the relation between NFC, ATI and CSE and AUP behavior, using a newly developed method that allows to measure individual differences in AUP behavior. In doing this, the first aim was to replicate the original study in which the AUP behavior method was developed to see if we could overcome some weaknesses and find the same results. Furthermore, we've studied the relation between individual differences such as NFC, ATI and CSE and differences in AUP behavior. Lastly, we've tried to develop a theory for the measurement of CSE (a dynamic trait) in relation to AUP behavior. Regarding the first aim, we managed to replicate the study with several adjustments to the method, but unfortunately stumbled upon some of the same measurement and methodological issues as the original research, which led us to conclude that the method is not yet optimal (i.e. valid) for measuring the full scope of AUP behavior. In light of that, it might not be surprising that we did not find support for our hypothesis that NFC, ATI and CSE are factors that can account for AUP resisting behaviors. Although we might have found these results because there is in fact no effect, we highly doubt that this is the case. We attribute the found results to measurement and methodology problems and therefore conclude that future research should focus on optimizing the operationalization of the method before conclusions can be drawn about the relation between AUP behavior and other constructs.

One major finding that led to our conclusion, is that the factor analysis did not fit our theoretical assumptions and expectations of AUP resistance behavior and might therefore -unintendedly- measure another construct. We were not able to calculate one AUP resistance measure, because the AUP measures did not seem to translate to the two underlying components of AUP behavior: invested cognitive effort and exploratory behavior. While we conclude that exploratory behavior was measured correctly, substantiated by the outcome of our expectations based on theory and the replication of the results of the research by Keil et al (2015), it seems we have not managed to measure invested cognitive effort. The component that we measured was related to using methods that are simple in use (low complexity and parameter demands), but also to using a lot of different methods and using methods that are very specific (i.e. only suitable for a specific task). Because Keil et al. (2015) suggested this component might be interpreted as knowledge transfer from previous experience, we compared the scores on this component to the frequency of previous experience with graphical programs and the participants own estimation of knowledge about graphical programs, but we found no relation, which makes the assumption that this component measures knowledge transfer less convincing. While we did find NFC to be a predictor of this component and ATI and CSE not to be, the interpretation of these results is ambiguous because we are not able to properly define the found component based on this study. In other words, if we have doubts about the underlying construct that we've measured (invested cognitive effort or knowledge transfer) and therefore with the validity of the measure, we ought to be cautiously with our conclusions about existing or non-existing relations between constructs. Because successful knowledge transfer and invested cognitive effort might be difficult constructs to distinguish in this experimental set-up, therefore creating the potential threat of a confound, we recommend creating a task that excludes the possibility for knowledge transfer. This would mean that the software used by the participants would have to be created so that it will have no to minimal overlap with other programs.

Table 8

Comparison of the factor loadings of the AUP behavior measures on the two extracted components (comp 1) cognitive effort/knowledge transfer and (comp 2) exploratory behavior between the current study and the study of Keil et al. (2015).

Current study	Comp 1	Comp 2	Keil et al.(2015)	Comp 1	Comp 2
Complexity	-.872			-.902	
Specifity	.868			.875	
Parameter Demands	-.840			-.838	
Method diversity	.722			.846	
Set parameter		.760			.661
Difficulty	.394	.653		.718	
RH duration		.624			.832
Undo Amount		.546			.886

Regarding the explorative behavior measure, we are more certain in our conclusion about the validity of the measures. We managed to largely replicate the findings of Keil et al. (2015), with a few exceptions. First, we found the measure 'difficulty' which to be a suitable measure for explorative behavior, in contrast to Keil et al. (2015), who did not find the measure to be related to the construct. While 'difficulty' was expected to be one of the measures for cognitive effort, there might be a simple explanation for this finding: individuals who explore more functions within a program, will probably discover more difficult tools and procedures. Secondly, under the assumption that the second component would measure explorative behavior as part of AUP resistance behavior, we hypothesized ATI, NFC and CSE to predict AUP resistance, and therefore explorative, behavior. Unfortunately, none of the personality traits seem to be related to explorative behavior. Again, this is in stark contrast to the findings of Keil et al. (2015), since they found a significant correlation of 0.5 between explorative behavior and NFC.

As a final step in the replication,we wanted to see if individuals who show more AUP resistance behavior are indeed more effective and efficient. Keil et al (2015) found time on task to be the only measure for effectiveness (effectiveness scores) and efficiency (time on task and method amount) that related to the explorative behavior measure, while we only found a significant relation between method amount and explorative behavior. Although we both studies found different results, both results lead to the same conclusion, namely that individuals who show more AUP resistance behavior are not more effective or efficient in our experiment. We add 'our experiment' to the conclusion, because one explanation for the unexpected results could be that there might not be enough repetition in our experiment for individuals to benefit from the time and cognitive effort invested in exploring and learning new methods. *If* participants took the time to read the hand-out or try-out different functions, the found functions might lead to a level of increased efficiency or effectiveness that is not sufficient to make up for the for the invested time or immediately lead to less faults (and therefore more effectiveness). Simply put, if one of our repetitive tasks takes 15 minutes to complete by using inefficient procedures and it takes 5 minutes to explore more effective and efficient procedures that will decrease the completion time to 10 minutes, we will not see an increase in efficiency in our experiment for those individuals. Therefore, future research could focus on increasing the amount of repetitions to such an extent that participants could in theory actually be more effective and efficient within the task. On the other hand, it could be stated that the tasks already included much repetition (as some participants informally let us know), since participants had to perform five versions of the same task in which each version consisted of repeatedly performing the same procedure. Raising the amount of repetitions in the experiment would decrease the external validity and would likely have a negative effect on participant's mood or affect, which could in turn have confounding effects on cognitive effort and explorative behavior. For example, research has shown that negative affect impairs cognitive performance by depleting working memory resources (Brose et al., 2012; as cited in Liew & Tan, 2016) and has a negative effect on problem solving performance (Baars, Wijnia & Paas, 2017). Therefore, our proposition for further research is to focus on other ways to measure effectiveness and efficiency. These might include measuring the increase in these constructs per individual per subtask (instead of per task) or extrapolating the scores. Most importantly, based on informal talks with and feedback from participants after the tasks, we would not recommend increasing the repetitions because we fear this might result in stress or negative affect.

The aforementioned differences between the original study and ours are peculiar, because we decided to only include task 1 and 3 in the analyses, which were the exact (unmodified) tasks as used by Keil et al (2015). Because of these opposing results, it seems preliminary to draw conclusions regarding AUP and relating individual differences. Rather, the methodology of the present study needs to be held to the light. There are a few suggestions to future studies in this line of research.

A proposal for further research, is to create a valid control measure for knowledge transfer. The measurement of cognitive effort/knowledge transfer seems to be a critical point for the usefulness of the developed AUP resistance method. While we anticipated this by creating a questionnaire that would allow us to check if the first component is indeed likely to measure knowledge transfer, we question the validity of the questionnaire. By means of this questionnaire, we tried to get information on the frequency and valence of usage of graphical programs. Therefore, we asked participants to indicate how often they had used several graphical programs thus far (never/seldom/monthly or more) and what their experience with these programs was (negative, slightly negative, slightly positive, positive,). Lastly, we asked participants to assess their knowledge of graphical programs on a scale from 1 (no knowledge) to 10 (high knowledge). Since previous research had shown that it's possible that the developed measure by Keil et al. (2015) actually is measuring knowledge transfer as one of the components, we wanted to include this questionnaire to check if this is hypothesis is likely. In retrospect, the questionnaire might say something about the previous experience and the perceived knowledge of graphical programs of the participant, but we question if we can infer something about knowledge *transfer* specifically.The amount of (perceived) knowledge an individual has, might not necessarily relate to the amount of knowledge transfer (i.e. someone with little knowledge about graphical programs might try to heavily rely on that little knowledge that person has, while someone with extensive knowledge might address the situation as a blank slate). Furthermore, the "objective" measure for knowledge, namely how often the participant has used a program, does actually noting more than that: indicate the frequency of the experience. It is unclear what can be inferred about knowledge or knowledge transfer. In spite of the aforementioned reasons, we used the questionnaire(s) to conclude that there is no relation between the found first component (invested cognitive effort/knowledge transfer) and knowledge and that it is therefore probably not measuring knowledge transfer. Since this conclusion might influence further research, we want to explicitly point out that we question the validity of this questionnaire and therefore this conclusion. For further research, we recommend to make it one of the main focuses in the study to investigate if the first component is actually measuring knowledge transfer (or not), by doing an extensive literature search on the construct and creating a valid and reliable measure to come to substantiated conclusion.

5.2 (Measuring) CSE

In line with our hypothesis, we have found that CSE does indeed change during tasks, and that this change is influenced by how successful individuals feel they've accomplished the task and by how difficult they perceived the task to be when their CSE beliefs before the task are (relatively) low. In contrast, when individuals have a *high* belief in their computer abilities (CSE), the perceived success does not affect their beliefs after the task and the perceived difficulty only has a small negative effect on their beliefs: when these individuals perceive the task to be difficult, this slightly decreases their beliefs in their abilities. While we did hypothesize that individuals with low CSE beliefs before the start of the task would be influenced by the perceived task success (where higher perceived task success would result in higher CSE beliefs after the task compared to lower perceived task success), we did not expect higher perceived difficulty to have a more positive effect on the CSE beliefs after the task than low perceived task difficulty. In other words, we believed that perceiving the task as difficult, would impair the beliefs in one's own abilities ('this is so hard to do, I must be less able than I thought'). What we found, in contrast, was that individuals seem to value it more (have more faith in their abilities) when a task went successful when the task was difficult, than when it was “easy”. Although this was not in concurrence with our expectations, this might be quite intuitive: when you complete something that is hard to do, you feel more confident or proud after completing that something then when it was not difficult at all. Our results show that this works the other way around as well: when individuals “fail” (had an unsuccessful experience), this affects their beliefs less when a task was perceived to be difficult (‘I didn’t manage, but in my offence: it was a difficult task’) then when a task was perceived as less difficult (‘I didn’t manage, but I should have since the task was not even that hard’). The results show that CSE levels do indeed change during relatively short tasks and experiment, which confirms our concerns about the measurement of CSE during experimental tasks and the AUP method in particular.

Another way the results can be viewed is in terms of saturation effects. As can be seen in table 7, both pre-CSE and perceived task success have a positive effect on post-CSE. However, the interaction of the two has a negative effect on post-CSE. This means that when individuals have either a high pre-CSE or high perceived task success this will increase post-CSE scores. But, having both will be less than simply the sum of both effects (hence the negative effect of the interaction). Just as with a typical learning curve (think for example the Rescorla Wagner model), the slope of this effect is not linear, but tends to flatten when getting closer to the maximum boundary. This can be interpret as a saturation effect : ‘the more of a similar is given, the closer it gets to the natural boundaries and the less it adds’ (Schmettow, 2020). In this case, pre-CSE and perceived task success might actually be part of a more or less similar cognitive process and therefore one might not ‘add much’ to the effect of the other when combined: the effect of the shared cognitive process was already doing its part. While we do have to be careful with drawing preliminary conclusions based on a model in which not all predictors reached statistical significance, it might be interesting to further explore these conditional effects in future research, because of the practical implications it might have. In the future, if one aims at creating an intervention for increasing post-CSE (for example when future research finds that it does help people to overcome falling for AUP behavior), the possible saturation effects might help guide the focus of that intervention. While this might still be a long way off, the findings are relevant for the short term as well, since it raises the question if there are other variables that are less similar to pre-CSE while having a greater effect on the dynamic nature of CSE.

We have to note that, in hindsight, we question the validity of several of the used questionnaires in this analysis. If we did not succeed in measuring the intended underlying constructs, this endangers the interpretability of the results. This concerns the CSE questionnaire and the task success questionnaire. First, while we constructed the questionnaires to the best of our knowledge, we did not perform an in-depth literature search to come to an exact definition of the construct and underlying factors for task success and task difficulty. For example, in constructing the perceived task success questionnaire, we assumed task success to have 3 underlying factors: success, efficiency and individual approach. The individual approach items focused on the strategy participants used, such as 'I've learned from the tasks' and 'I've tried to show my skills during the tasks'. While it is true that psychological theories emphasize the difference in individuals in their goal orientation (learning goal vs performance goal) and what outcome they perceive to be successful (learning vs showing ability) (Yi & Hwang ,2003), our questionnaire is not able to differentiate between these individuals because it uses a sum-score of all these items. In other words, participants who learned *and* showed their ability were seen as having a higher perceived task success than participants who had only one of these outcomes, while the latter may be perfectly fine with just one of these outcomes because of their goal orientation and therefore perceive their experience as just as successful as the former group.

Furthermore, there are some factors that indicate there might be some issues with the validity of the CSE questionnaire as well. While our questionnaire was based on the validated questionnaire of Murphy, Coover, and Owen (1989), we made several adjustments to the questionnaire which might have unintendedly have an effect on the psychometric qualities. One of these (necessary) adjustments is that we reformulated the items to fit contemporary times, because the way we interact with technology has rapidly changed since the validation of the original questionnaire. The original questionnaire consisted of items such as 'I feel confident entering and saving data (numbers or text) in a file', 'handling a floppy disc correctly' and 'logging onto a mainframe computer system'. We excluded some items (such as the items concerning the mainframe), altered some items to modern terms (such as 'I feel confident using a USB-stick') and reused items that we deemed sufficient and still relevant (such as 'I feel confident entering and saving data (numbers or text) in a file'). In doing this, we might have had a too conservative approach, causing ceiling effects. This is reflected in the data by high mean scores on the pre-CSE and post-CSE questionnaires and the low variance (M=5.24, SD=.93 and M=5.23, SD=.97, respectively. Since our sample was quite divers in terms of professions, including individuals who do not mainly work with computers, such as a cook, make-up artist, tour guide or self-employed psychologist (in contrast to studies with only students) and in age (20 – 35 years old), having a homogeneous group cannot be the reason for high mean scores and the low variance. Consequently, we fear that the updated questionnaire might not be able to sufficiently discriminate between CSE levels in our sample and the population. A possible reason that a (relatively conservative) update of a 1989 rating scale might show ceiling effects in a 2018 study, might have to do with the response processes of participants. In particular with anchoring: the defining of the extremes by participants to create a meaningful range to which a participant can compare the level of their feelings. While non-experts in 1989 might have had several interactions with computers, the amount and scope of these experiences were far less in comparison with our contemporary computer use and experiences. Our sample, 18 – 35 year old participants, have interacted regularly with computers since their childhood or teenage years. When asking a participant in 1989 to think about a situation in which they felt uncertain in their computer abilities (lowest extreme), they might remember the time they wanted to make a back-up or organize their computer. If you ask that same question in 2018, participants are far more likely to think of more advanced and specific or situations, such as using Rstudio, Indesign or Matlab, because practically all participants have learned and used the 'general' computer skills growing up. Therefore, the question rises if asking about the level of confidence in general computer skills in 2018 can result in a response of 'not at all' *at all* and if response scale options should be revised so that it is possible to differentiate on the upper side of the dimension of CSE.

In sum, the described problems with the two questionnaires impact our study in the sense that we have to be cautious with our interpretations and conclusions based on these questionnaires. It does seem like there is an interesting interplay between CSE, perceived task difficulty and perceived task success, but further research is necessary to see if the found effects are still present when other (or adjusted) questionnaires are used. We expect that the valence of the effects to remain unchanged, but that the strength of the effects could increase. Further research is necessary, because simply measuring a dynamic trait at one point during the AUP resistance method can cause bias in the results: there might be a different relation between pre-CSE measures and AUP behavior than post-CSE levels and AUP behavior, because pre-CSE and post-CSE levels differ. This study forms a basis for further research into how to include dynamic traits in the AUP resistance measures with respect to their dynamic nature.

To conclude: this study has made some important further steps in the development of a method to measure individual differences in AUP behavior. Such a method would make it possible to measure individual differences in AUP behavior, creating opportunities for researchers and developers to propose theories which can form a basis for creating tools or interventions for specific groups of users, or even tailored to the individual user. Our research supports previous research in that it shows that measuring AUP behavior is complex, and that creating a valid and reliable measure is difficult, but achievable. One of the two underlying factors of AUP resistance behavior is explorative behavior, which has been successfully measured in this study as well as in a previous study. The second underlying factor of AUP resistance behavior, invested cognitive effort, seems to be more difficult to measure. The current scoring of the method does not seem suitable to measure invested cognitive effort yet, but several proposals for further research could help develop the method so that it will be in the future. While we used the method to see if individual differences such as NFC, ATI and CSE can account for differences in AUP resistance behavior, we want to emphasize that the found effects (or the lack thereof), should be interpreted carefully: if the method is not yet suitable to measure the full definition of AUP behavior, we cannot conclude that characteristics do or do not relate to this full definition of AUP behavior. Found null-effects, such as in the current study, could then potentially steer the research field in the wrong direction. Therefore, our most important recommendation is to first optimize the method, before proceeding with theorization about the relation between AUP behavior and specific personality traits and characteristics. Concerning this last step, our study has shown that special attention should be given to individual differences that have a dynamic nature and influenced by the task itself (i.e. dynamic traits or states). Both theoretical and methodological, we are still in search of the proper way of capturing this dynamic nature in our attempt to study the relation between the characteristic (in our case, CSE) and AUP behavior. We have proposed a working theory which can serve as a starting point for further research in CSE specifically, and dynamic traits in general.

In the future, the further optimization of the AUP method could contribute to the opportunity to -at last- gain insight into how or why users differ in their invested effort and the exploration of new functions, allowing them to be more effective and efficient. Answers to these questions are becoming more and more important in this age in which effectiveness and efficiency are valued highly and technologies are expected to ‘fit’ the user (i.e. be able to differentiate between users or adapt to the individual user and his or her needs). We as HCI researchers should invest the necessary cognitive effort and turn on our explorative minds to find new methods and answers to these questions and resist the tendency to hold on to the familiar and our old ways.

6. References

Baars, M., Wijnia, L., & Paas, F. (2017). The association between motivation, affect, and self-regulated learning when solving problems. *Frontiers in psychology*, *8*, 1346. https://doi.org/10.3389/fpsyg.2017.01346

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, *50*(2), 248-287. https://doi.org/10.1016/0749-5978(91)90022-L

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191. https://doi.org/10.1016/0146-6402(78)90002-4

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of personality assessment*, *48*(3), 306-307. https://doi.org/10.1207/s15327752jpa4803_13

Carroll, J. M., & Rosson, M. B. (1987). Paradox of the Active USer. *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. https://doi.org/10.1017/CBO9781107415324.004

Corbett, E., & Weber, A. (2016). What can i say? Addressing user experience challenges of a mobile voice user interface for accessibility. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016*, 72–82. https://doi.org/10.1145/2935334.2935386

Franke, T., Attig, C., & Wessel, D. (2017). Assessing Affinity for Technology Interaction–The Affinity for Technology Interaction (ATI) Scale. *Unpublished manuscript*.

FrØkjaer, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? *Conference on Human Factors in Computing Systems - Proceedings*, *2*(1), 345–352. https://doi.org/10.1145/332040.332455

Fu, W. T., & Gray, W. D. (2004). Resolving the paradox of the active user: Stable subuptional performance in interactive tasks. *Cognitive Science*. https://doi.org/10.1016/j.cogsci.2004.03.005

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies*, *64*(2), 79-102. https://doi.org/10.1016/j.ijhcs.2005.06.002

Keil, J., Schmettow, M., & Noordzij, M. (2015). Exploring Individual Differences in the Tendency to Fall For the Active User Paradox Table of contents.

Liew, T. W., & Tan, S. M. (2016). The effects of positive and negative mood on cognition and motivation in multimedia learning environment. *Journal of Educational Technology & society*, *19*(2), 104-115.

Marks, R., & Allegrante, J. P. (2005). A review and synthesis of research evidence for self-efficacy-enhancing interventions for reducing chronic disability: implications for health education practice (part II). *Health promotion practice*, *6*(2), 148-156. https://doi.org/10.1177/1524839904266792

Murphy, C. A., Coover, D., & Owen, S. V. (1989). Development and Validation of the Computer

Self-Efficacy Scale. *Educational and Psychological Measurement*. https://doi.org/10.1177/001316448904900412

Niksirat, K. S., Silpasuwanchai, C., Mohamed, M., Ahmed, H., Cheng, P., & Ren, X. (n.d.). A Framework for Interactive Mindfulness Meditation Using Attention-Regulation Process. https://doi.org/10.1145/3025453.3025914

Sadowski, C. J. (1993). An Examination of the Short Need for Cognition Scale An Examination of the Short Need for Cognition Scale. *The Journal of Psychology*, *3980*(127:4), 451–454. https://doi.org/10.1080/00223980.1993.9915581

Schmettow, M. (2020). *New Statistics for the design researcher.* Unpublished manuscript.

Schmettow, M., Noordzij, M. L., & Mundt, M. (2013). An implicit test of UX: individuals differ in what they associate with computers. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2039-2048). https://doi.org/10.1145/2468356.2468722

Schunk, D. H., Erttner, P. A., Ahn, J., Chen, J., Gentry, A., Hruskocy, C., … Young, J. (1999). Self-Regulatory Processes During Computer Skill Acquisition: Goal and Self-Evaluative Influences. *Journal of Educational Psychology*, *91*(2), 251–260.

Yi, M. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: Self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human Computer Studies*. https://doi.org/10.1016/S1071-5819(03)00114-9

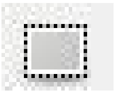
7. Appendix

7.1 GIMP tool guide

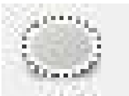
GIMP Gereedschappen-handleiding

Op de volgende paginas worden de meest belangrijke gereedschappen van GIMP uitgelegd. Je mag deze handleiding altijd gebruiken om naar functies te zoeken, die je bij de uitvoer van de opgaven kunnen helpen.

Selectie-gereedschappen



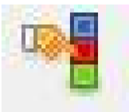
Rechthoekige selectie: Selecteert een rechthoekig gebied.



Ovale selectie: Selecteert een ovale gebied.Om de selectie rond te maken, o nder Gereedschapsopties *Vast: Verhouding*activeren.

Vrije selectie: Vrije vormen s electeren. Om een bereik te selecteren linke mu isknop drukken en een vorm met de muis tekenen. Zodra de vorm geloten is (gele rondje verschijnt bij de muisaanwijzer) linke muisknop loslaten.

Toverstaf: Kiest een samenhangende bereik met gelijksoortige kleuren. Om te selecteren, de gewenste bereik met de linke muisknop aanklikken. Hoe hoger de *Drempelwaarde* in de Gereedschapsopties is, des te hoger is de tolerantie van de selectie. (kleuren worden deel van de selectie die iets van de aangeklikte kleur afwijken)



Selecteren op kleur: Wordt een gekleurd bereik in deafbeelding aangeklikt, d an worden alle bereiken in de afbeelding met soortgelijke k leuren geselecteerd,Hoe hoger de *Drempelwaa rde* in de Gereedschapsopties is, des te hoger is de toleran tie van de selectie. (kleuren worden deel van de selectie die iets van de aangeklikte kleur afwijken)

Tips:

- Om de actuele selectie om teker n kies: *Selecteren>Inverteren*
- Is een bereik geselecteerd, dan w erken teken-gereedschappen alleen in dit bereik. Om de selectie op te heffen kies: *Selectie>Niets*
- Om meerdere bereiken tegelijk te selecteren moet je bij het selecteren de SHIFT-toets drukken.
- DEL (ENTF)-toets drukken of *Bewerken>Knippen* selecteren om de inhoud van een s electie te verwijderen

Teken-gereedschappen

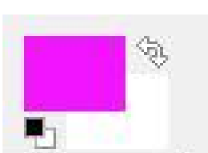
Emmer: Op een bereik met een bepaalde kleur klikken om deze met de gekozen gereedschap-voorggrondkleur te vullen. Hierbij moet in de Gereedschapsopties *Gelijke kleuren vullen* geactiveerd zijn. Hoe hoger de *Drempelwaarde* in de Gereedschapsopties is, des te meer kleuren worden gevuld die iets van de gekozen kleur-bereik afwijken.
Om een gehele selectie te vullen moet je in de Gereedschapsopties *Hele selectie vullen* activeren en op de selectie klikken.

Potlood: Linke muisknop drukken om vrij met de gekozen gereedschap-voorgr ondkleur op de afbeelding te schilderen. De *grootte* van he t potlood kan in de Gereedschapsopties aangepa st worden. Hoe hoger de *Dekking* in de Gereedschapso pties, des te minder doorschijnend is de geschilderde bereik. (Is de *Dekking* bij 100, dan verdwijnen de oude kleuren helemaal)

Penseel: Linke muisknop druk ken om vrij met de gekozen gereedschap-voorgrondkleur op de afbeelding te schilderen. De *grootte* van het potlood kan in de Gereedschapsopties aangepast worden. Hoe hoger de *Dekking* in de Gereedschapso pties, des te minder doorschijnend is de geschilderde bereik. Anders dan bij het potlood heeft het penseel gee n gelijkmatige dekking.De dekking is aan de kan ten minder sterk.

Gum: Linke muisknop drukken om bereiken van de afbeelding te verwijderen. Hoe hoger de *Dekking* in de Gereedschapsopties, des te meer kleur wordt met een klik verwijderd. (Is de Dekking op 100, dan worden alle kleuren in het gekozen bereik met een klik verwijderd)

Tips:



Om de gereedschap-voorggrondkleur te veranderen moet je op het voorgrondkleur veld (beneden in pink) klikken.

Andere gereedschappen

Pipet: Op een kleur in de afbeelding klikken om deze als gereedschap-voorgrondkleur te selecteren.

Vergrootglas: Op een bereik van de afbeelding klikken om deze te vergroten. In de gereedschapsopties kan je tussen vergroten (*Zoom in*) en verkleinen (*Zoom out*) kiezen.

7.2 Gimp setup + task instruction

Taak 1 instructies (GIMP)

Op het bureaublad vind je de gereedschappenhandleiding die je mag raadplegen. Maak gebruik van de muis, en niet van short-cuts op je toetsenbord (bijvoorbeeld 'cmd+Z').

1)

- Open het bestand **task 1 v01**
- Verwijder alle **rode** objecten in de afbeelding, probeer hierbij andere objecten min mogelijk te beschadigen of aan te passen
- Sla het bestand op (bureaublad)

2)

- Open het bestand **task 1 v02**
- Verwijder alle **groene** objecten in de afbeelding, probeer hierbij andere objecten zo min mogelijk te beschadigen of aan te passen
- Sla het bestand op

3)

- Open het bestand **task 1 v03**
- Verwijder alle **violette** objecten in de afbeelding, probeer hierbij andere objecten zo min mogelijk te beschadigen of aan te passen
- Sla het bestand op

4)

- Open het bestand **task 1 v04**
- Verwijder alle **blauwe** objecten in de afbeelding, probeer hierbij andere objecten zo min mogelijk te beschadigen of aan te passen
- Sla het bestand op

5)

- Open het bestand **task 1 v05**
- Verwijder alle **groene** objecten in de afbeelding, probeer hierbij andere objecten zo min mogelijk te beschadigen of aan te passen
- Sla het bestand op

Einde van de eerste taak. Ga door met taak 2 op de volgende pagina.

Taak 2 instructies (GIMP)

1)

- Open het bestand **task 2 v01**
- Verander deze afbeelding naar een rode achtergrond met een witte cirkel (op dezelfde plaats als de huidige cirkel)
- Sla het bestand op (bureaublad)

2)

- Open het bestand **task 2 v02.**
- Verander deze afbeelding naar een blauwe achtergrond met een geel vierkant (op dezelfde plek als het huidige vierkant)
- Sla het bestand op.

3)

- Open het bestand **task 2 v03.**
- Verander deze afbeelding naar een gele achtergrond met een blauwe ster (op dezelfde plek als de huidige ster)
- Sla het bestand op

4)

- Open het bestand **task 2 v04.**
- Verander deze afbeelding naar een blauwe achtergrond met een rood rechthoek (op dezelfde plek als de huidige rechthoek)
- Sla het bestand op

5)

- Open het bestand **task 2 v05.**
- Verander deze afbeelding naar een witte achtergrond met een gele circel (op dezelfde plek als de huidige rechthoek)
- Sla het bestand op.

Einde van de tweede taak. Ga door met de laatste taak op de volgende pagina.

Taak 3 instructies (GIMP)

1)

- Open het bestand **task 3 v01**
- Verwijder alle streepjes in de afbeelding. Probeer hierbij de rode rondjes en sterren zo weinig mogelijk te beschadigen
- Sla het bestand op

2)

- Open het bestand **task 3 v02**
- Verwijder alle streepjes in de afbeelding. Probeer hierbij de blauwe rondjes en sterren zo weinig mogelijk te beschadigen
- Sla het bestand op

3)

- Open het bestand **task 3 v03**
- Verwijder alle streepjes in de afbeelding. Probeer hierbij de groene rondjes en sterren zo weinig mogelijk te beschadigen
- Sla het bestand op

4)

- Open het bestand **task 3 v04**
- Verwijder alle streepjes in de afbeelding. Probeer hierbij de gele rondjes en sterren zo weinig mogelijk te beschadigen
- Sla het bestand op

5)

- Open het bestand **task 3 v05**
- Verwijder alle streepjes in de afbeelding. Probeer hierbij de oranje rondjes en sterren zo weinig mogelijk te beschadigen
- Sla het bestand op

Einde van de laatste taak. Vul nu de overige vragenlijsten in. Hierna kun je de onderzoeker er weer bij halen.

7.3 Informed consent

GEÏNFORMEERDE TOESTEMMING

Ik, ………………………………………………………………….. (*naam proefpersoon)*

stem toe mee te doen aan een onderzoek dat uitgevoerd wordt door

R.D Timmer

Ik ben me ervan bewust dat deelname aan dit onderzoek geheel vrijwillig is. Ik kan mijn medewerking op elk tijdstip stopzetten en de gegevens verkregen uit dit onderzoek terugkrijgen, laten verwijderen uit de database, of laten vernietigen.

De volgende punten zijn aan mij uitgelegd:

1. Het doel van dit onderzoek is inzicht te krijgen in hoe interfaces van programma's gebruikt kunnen worden
2. Er zal mmij gevraagd worden verschillende vragenlijsten in te vullen en verschillende taken met het programma GIMP uit te voeren.
3. Het gehele onderzoek zal ongeveer 45 minuten duren. Aan het einde van het onderzoek zal de onderzoeker uitleggen waar het onderzoek over ging.
4. De gegevens verkregen uit dit onderzoek zullen anoniem verwerkt worden en kunnen daarom niet bekend gemaakt worden op een individueel identificeerbare manier.
5. Er behoort geen stress of ongemak voort te vloeien uit deelname aan dit onderzoek.
6. De onderzoeker zal alle verdere vragen over dit onderzoek beantwoorden, nu of gedurende het verdere verloop van het onderzoek.

Voor eventuele klachten over dit onderzoek kunt u zich wenden tot de secretaris van de Commissie Ethiek van de faculteit Gedragswetenschappen van de Universiteit Twente, mevr. J. Rademaker (telefoon: 053-4894591; e-mail:j.rademaker@utwente.nl, Postbus 217, 7500 AE Enschede).

Handtekening onderzoeker: ……………………………………
Datum: …………………..

Handtekening proefpersoon: ……………………………………
Datum: …………………..

7.4 Demographic questionnaire

Demografische vragenlijst (1)

1. Geslacht:

   Leeftijd:

   Beroep/studie:

   Hoe vaak gebruik je een computer (bijv. dagelijks/een paar keer per week/maand):

2. Kruis aan hoe vaak je de volgende programma's tot nu toe gebruikt hebt

	Nooit	**Zelden**	**Maandelijks of vaker**
Microsoft Paint	O	O	O
Adobe Photoshop	O	O	O
MacPaint	O	O	O
GIMP	O	O	O
Paintbrush	O	O	O

3. Kruis aan wat je eerdere ervaringen zijn met onderstaande programma's. Mocht je geen ervaring met een programma hebben, laat deze dan open.

	Negatief	**Overwegend negatief**	**Overwegend positief**	**Positief**
Microsoft Paint	O	O	O	O
Adobe Photoshop	O	O	O	O
MacPaint	O	O	O	O
GIMP	O	O	O	O
Paintbrush	O	O	O	O

4. Hoe beoordeel je jouw kennis van grafische programma’s op een schaal van 0 (helemaal geen voorkennis) tot 10 (heel veel kennis)?

…………………………………………………………

7.5 Affinity for Technology Interaction (ATI) Scale

The original English version of the ATI (Franke, Attig, & Wessel, 2018) was translated to Dutch, which resulted in the following version:

Vragenlijst Technische systemen (2)

In deze vragenlijst zullen wij je vragen naar je interactie met technische systemen. De term 'technische systemen' refereert zowel naar apps en andere softwareapplicaties, als naar volledige digitale apparaten (bijv. mobile telefoons, computers, Tv's, navigatie in de auto)

Geef aan in welke mate je het eens of oneens bent met de volgende stellingen	Volledig mee oneens	Grotendeels mee oneens	Deels mee oneens	Deels mee eens	Grotendeels mee eens	Volledig mee eens
1. Ik vind het leuk om tot in detail bezig te zijn met technische systemen.	○	○	○	○	○	○
2. Ik vind het leuk functies van nieuwe technische systemen te testen.	○	○	○	○	○	○
3. Ik ga voornamelijk om met technische systemen omdat het moet.	○	○	○	○	○	○
4. Wanneer ik een nieuw technisch systeem voor mij heb, probeer ik alles uit.	○	○	○	○	○	○
5. Ik spendeer graag tijd aan het bekend raken met een nieuw technisch systeem.	○	○	○	○	○	○
6. Het is voor mij voldoende dat een technisch systeem werkt; het maakt mij niet uit hoe of waarom.	○	○	○	○	○	○
7. Ik probeer te begrijpen hoe een technisch systeem precies werkt.	○	○	○	○	○	○
8. Het is voor mij voldoende om de basisfuncties van een technisch systeem te kennen.	○	○	○	○	○	○
9. Ik probeer volledig gebruik te maken van alle mogelijkheden van een technisch systeem.	○	○	○	○	○	○

7.6 Need For Cognition scale (NFC)

Vragenlijst denken (3)

In hoeverre zijn de onderstaande uitspraken op jou van toepassing?

	Helemaal niet						**Helemaal wel**
1. Als ik moet kiezen heb ik liever een ingewikkeld dan een simpel probleem	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
2. Ik ben graag verantwoordelijk voor een situatie waarin veel nagedacht moet worden	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
3. Nadenken doe ik niet voor m'n plezier	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
4. Ik doe liever iets waarbij weinig nagedacht hoeft te worden dan iets waarbij mijn denkvermogen zeker op de proef wordt gesteld	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
5. Ik probeer situaties te vermijden waarin de kans groot is dat ik diep over iets moet nadenken	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
6. Iets langdurig en nauwgezet afwegen geeft mij voldoening.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
7. Ik denk alleen zoveel als nodig is.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
8. Ik denk liever na over kleine dagelijkse dingen dan over lange-termijn zaken.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O

9. Ik hou van taken waarbij weinig nagedacht hoeft te worden als ik ze eenmaal geleerd heb.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
10. Het idee dat je op je verstand moet vertrouwen om top te bereiken spreekt mij aan.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
11. Ik geniet echt van een taak waarbij je met nieuwe oplossingen voor problemen moet komen.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
12. Het leren van nieuwe manieren om te denken vind ik niet erg boeiend.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
13. Ik vind het prettig als mijn leven gevuld is met puzzels die ik moet oplossen.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
14. Abstract denken is een bezigheid die mij aanspreekt.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
15. Ik heb liever een taak die intellectueel, moeilijk en belangrijk is, dan een taak die enigszins belangrijk is, maar waarbij je niet veel hoeft na te denken.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
16. Als ik een taak heb voltooid de veel inspanning heeft gevergd ben ik eerder opgelucht dan voldaan.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
17. Ik vind het voldoende wanneer iets blijkt te werken: hoe of waarom het precies werkt interesseert mij niet.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
18. Gewoonlijk denk ik uitgebreid na over zaken, zelfs wanneer ze mij niet persoonlijk aangaan.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O

7.7 Computer Self-Efficacy (CSE) questionnaire

The original CSE questionnaire (xxx) was updated and translated into Dutch.

In hoeverre zijn de onderstaande uitspraken <u>op dit moment</u> op jou van toepassing?

	Helemaal niet			**neutraal**			**Helemaal wel**
1. Ik voel mij zelfverzekerd in het maken van een backup van mijn volledige werk	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
2. Ik voel mij zelfverzekerd in het succesvol gebruiken van een USB-stick.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
3. Ik voel mij zelfverzekerd in het vinden van hulp op de juiste plek wanneer het nodig is.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
4. Ik voel mij zelfverzekerd in het leren van gevorderde vaardigheden binnen een specifiek programma (software)	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
5. Ik voel mij zelfverzekerd in het openen van bestanden met het juiste programma.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
6. Ik voel mij zelfverzekerd in het begrijpen van woorden/termen gerelateerd aan computer *hardware.*	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
7. Ik voel mij zelfverzekerd in het begrijpen van woorden/termen gerelateerd aan computer *software*.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
8. Ik voel mij zelfverzekerd in het installeren van software en het aan de praat krijgen.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
9. Ik voel mij zelfverzekerd in het leren gebruiken van verschillende programma's (software).	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
10. Ik voel mij zelfverzekerd in het invoeren en het bewaren van data (nummers en woorden) in een bestand.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
	1.	**2**	**3**	**4**	**5**	**6**	**7.**

11. Ik voel mij zelfverzekerd in het vinden van functies in het menu van een software.	O	O	O	O	O	O	O
12. Ik voel mij zelfverzekerd in het analyseren van numerieke data op mijn computer.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
13. Ik voel mij zelfverzekerd in het werken op een computer.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
14. Ik voel mij zelfverzekerd in het organiseren van informatie op de computer	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
15. Ik voel mij zelfverzekerd in het toevoegen of veranderen van data in een tabellarisch data bestand (e.g. Exel, SPSS).	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
16. Ik voel mij zelfverzekerd in het uitleggen waarom een programma wel of niet werkt.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
17. Ik voel mij zelfverzekerd in het schrijven van een email of essay op de computer.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
18. Ik voel mij zelfverzekerd in het beschrijven van de functie van computer hardware	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
19. Ik voel mij zelfverzekerd in het begrijpen van de drie stappen van data verwerking: input, processing en output.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
20. Ik voel mij zelfverzekerd in het vinden van hulp voor problemen in het computer systeem.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
21. Ik voel mij zelfverzekerd in het schrijven van simpele code in een programmeertaal.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
22. Ik voel mij zelfverzekerd in het kopiëren van een bestand naar een andere locatie.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O
23. Ik voel mij zelfverzekerd in het verwijderen van bestanden wanneer ze niet langer nodig zijn.	**1.** O	**2** O	**3** O	**4** O	**5** O	**6** O	**7.** O

7.8 Perceived task success & task difficulty

Items to measure perceived task succes and perceived task difficulty where presented in a combined questionnaire. Item 12, 13, 14 and 15 measured task difficulty.

Taak reflectie vragenlijst (5)

Deze vragenlijst betreft jouw ervaring tijdens de taken. Geef aan in hoeverre je het eens of oneens bent met onderstaande stellingen.

	Helemaal niet mee eens	**Deels mee oneens**	**Neutraal**	**Deels mee eens**	**Helemaal mee eens**
1. Ik ben tevreden met de kwaliteit van mijn resultaten	**1.** O	**2** O	**3** O	**4** O	**5** O
2. Mijn resultaten zijn van hoge kwaliteit	**1.** O	**2** O	**3** O	**4** O	**5** O
3. Ik ben tevreden met mijn werkwijze gedurende de taken	**1.** O	**2** O	**3** O	**4** O	**5** O
4. Het koste mij weinig inspanning om de taken te volbrengen	**1.** O	**2** O	**3** O	**4** O	**5** O
5. De kwaliteit van mijn resultaten zijn hoog in vergelijking met de inspanning die ik heb geleverd	**1.** O	**2** O	**3** O	**4** O	**5** O
6. Ik heb de taken relatief snel afgerond	**1.** O	**2** O	**3** O	**4** O	**5** O
7. Ik heb geprobeerd te leren van de taken	**1.** O	**2** O	**3** O	**4** O	**5** O

8. Ik heb van de taken geleerd	**1.** O	**2** O	**3** O	**4** O	**5** O
9. Ik heb geprobeerd mijn vaardigheden te tonen tijdens het doen van de taak	**1.** O	**2** O	**3** O	**4** O	**5** O
10. Ik heb mijn vaardigheden kunnen tonen tijdens te taak.	**1.** O	**2** O	**3** O	**4** O	**5** O
11. Ik heb mijn best gedaan op de taken	**1.** O	**2** O	**3** O	**4** O	**5** O
12. De taken succesvol afronden was gemakkelijk	**1.** O	**2** O	**3** O	**4** O	**5** O
13. Ik heb de taken met weinig inspanning afgerond	**1.** O	**2** O	**3** O	**4** O	**5** O
14. De taken succesvol afronden was moeilijk	**1.** O	**2** O	**3** O	**4** O	**5** O
15. Ik heb de taken als moeilijk ervaren	**1.** O	**2** O	**3** O	**4** O	**5** O

7.9 SPSS syntax

7.9.1 Variance AUP measures

```
DATASET ACTIVATE Final_selection.
EXAMINE VARIABLES=Specifity__mean Difficulty__mean Complexity_mean
DelayedFeedback_mean
    ParameterDemands_mean method_divers Parameters_sum UndoAmount_sum
RHDuration_Sec_sum BY Task
  /PLOT BOXPLOT HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

DATASET ACTIVATE Final_selection.
DESCRIPTIVES VARIABLES=Specifity__mean Difficulty__mean Complexity_mean
ParameterDemands_mean
    method_divers Parameters_sum UndoAmount_sum RHDuration_Sec_sum
  /SAVE
  /STATISTICS=STDDEV.
```

7.9.2 Factor analyses AUP measures

→ Correlaties berekenen

```
DATASET ACTIVATE One_measure_final!

CORRELATIONS
  /VARIABLES=ZUndoAmount_mean ZRHDuration_Sec_mean ZSpecify__mean
ZDifficulty__mean
    ZComplexity_mean ZParameterDemands_mean Zmethod_divers_mean
ZParameters_mean
  /PRINT=TWOTAIL NOSIG
  /STATISTICS DESCRIPTIVES
  /MISSING=PAIRWISE.
```

- ➔ First factor analysis

```
FACTOR
  /VARIABLES ZSpecify__mean ZDifficulty__mean ZComplexity_mean
ZParameterDemands_mean
    Zmethod_divers_mean ZParameters_mean ZUndoAmount_mean
ZRHDuration_Sec_mean
  /MISSING LISTWISE
```

```
 /ANALYSIS ZSpecifity__mean ZDifficulty__mean ZComplexity_mean
ZParameterDemands_mean
   Zmethod_divers_mean ZParameters_mean ZUndoAmount_mean
ZRHDuration_Sec_mean
 /PRINT UNIVARIATE INITIAL CORRELATION SIG KMO AIC EXTRACTION
ROTATION
 /FORMAT BLANK(.10)
 /CRITERIA MINEIGEN(1) ITERATE(25)
 /EXTRACTION PC
 /CRITERIA ITERATE(25) DELTA(0)
 /ROTATION OBLIMIN
 /SAVE REG(ALL)
 /METHOD=CORRELATION.
```

-> Second factor analysis

```
FACTOR
 /VARIABLES ZSpecifity__mean ZDifficulty__mean ZComplexity_mean
ZParameterDemands_mean
   Zmethod_divers_mean ZParameters_mean ZUndoAmount_mean
ZRHDuration_Sec_mean
 /MISSING LISTWISE
 /ANALYSIS ZSpecifity__mean ZDifficulty__mean ZComplexity_mean
ZParameterDemands_mean
   Zmethod_divers_mean ZParameters_mean ZUndoAmount_mean
ZRHDuration_Sec_mean
 /PRINT UNIVARIATE INITIAL CORRELATION SIG KMO AIC EXTRACTION
ROTATION
 /FORMAT BLANK(.3)
 /PLOT EIGEN
 /CRITERIA FACTORS(2) ITERATE(25)
 /EXTRACTION PC
 /CRITERIA ITERATE(25) DELTA(0)
 /ROTATION OBLIMIN
 /SAVE REG(ALL)
 /METHOD=CORRELATION.
```

- ➔ Explorative: correlation graphical knowledge, previous experience and fact 1.

```
CORRELATIONS
 /VARIABLES=FAC1_1 GraphKnowlegde SUM_FREQ
 /PRINT=TWOTAIL NOSIG
 /STATISTICS DESCRIPTIVES
 /MISSING=PAIRWISE.
```

7.9.3 Efficiency and effectiveness

```
CORRELATIONS
  /VARIABLES=FAC1_1 FAC2_1 Effectiveness_mean Method_amount_mean
Time_on_task_mean
  /PRINT=TWOTAIL NOSIG
  /STATISTICS DESCRIPTIVES
  /MISSING=PAIRWISE.
```

7.9.4 NFC, ATI & (pre) CSE

```
CORRELATIONS
  /VARIABLES=FAC1_1 SUM_FREQ
  /PRINT=TWOTAIL NOSIG
  /STATISTICS DESCRIPTIVES
  /MISSING=PAIRWISE.

* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=FAC1_1 FAC2_1
MEAN_NFC MEAN_ATI MEAN_CSE_PRE
    MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE
  /FITLINE TOTAL=NO.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: FAC1_1=col(source(s), name("FAC1_1"))
  DATA: FAC2_1=col(source(s), name("FAC2_1"))
  DATA: MEAN_NFC=col(source(s), name("MEAN_NFC"))
  DATA: MEAN_ATI=col(source(s), name("MEAN_ATI"))
  DATA: MEAN_CSE_PRE=col(source(s), name("MEAN_CSE_PRE"))
  GUIDE: axis(dim(1.1), ticks(null()))
  GUIDE: axis(dim(2.1), ticks(null()))
  GUIDE: axis(dim(1), gap(0px))
  GUIDE: axis(dim(2), gap(0px))
  GUIDE: text.title(label("Scatterplot Matrix REGR factor score   1 for analysis 1,REGR
factor ",
    "score   2 for analysis 1,MEAN_NFC..."))
  TRANS: FAC1_1_label = eval("REGR factor score   1 for analysis 1")
  TRANS: FAC2_1_label = eval("REGR factor score   2 for analysis 1")
  TRANS: MEAN_NFC_label = eval("MEAN_NFC")
  TRANS: MEAN_ATI_label = eval("MEAN_ATI")
  TRANS: MEAN_CSE_PRE_label = eval("MEAN_CSE_PRE")
```

```
 ELEMENT:
point(position((FAC1_1/FAC1_1_label+FAC2_1/FAC2_1_label+MEAN_NFC/MEAN_
NFC_label+

MEAN_ATI/MEAN_ATI_label+MEAN_CSE_PRE/MEAN_CSE_PRE_label)*(FAC1_
1/FAC1_1_label+FAC2_1/FAC2_1_label+

MEAN_NFC/MEAN_NFC_label+MEAN_ATI/MEAN_ATI_label+MEAN_CSE_PRE/
MEAN_CSE_PRE_label)))
END GPL.

REGRESSION
 /DESCRIPTIVES MEAN STDDEV CORR SIG N
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT FAC1_1
 /METHOD=ENTER MEAN_ATI
 /METHOD=ENTER MEAN_NFC
 /METHOD=ENTER MEAN_CSE_PRE
 /PARTIALPLOT ALL
 /SCATTERPLOT=(*ZRESID ,*ZPRED)
 /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
 /CASEWISE PLOT(ZRESID) OUTLIERS(3)

REGRESSION
 /DESCRIPTIVES MEAN STDDEV CORR SIG N
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS CI(95) BCOV R ANOVA COLLIN TOL CHANGE ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT FAC2_1
 /METHOD=ENTER MEAN_ATI
 /METHOD=ENTER MEAN_NFC
 /METHOD=ENTER MEAN_CSE_PRE
 /PARTIALPLOT ALL
 /SCATTERPLOT=(*ZRESID ,*ZPRED)
 /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
 /CASEWISE PLOT(ZRESID) OUTLIERS(3)
```

7.9.5 Exploratory analyses

Note: variable name of MEAN_CSE_PRE was renamed MEAN_PRE to run this analysis. Furthermore, no syntax is availeble, therefore the setting are specified below.

- ➔ Moderation using Process. Model 2.
- ➔ X = MEAN_PRE
  Y= MEAN_POST
  W=Mean_suc
  Z=Mean_dif

- ➔ Options:
  Covariance matric
  Generate code for visualization
  Residual correlations
  Centering: all variables that define products
  Conditioning values: -1SD, Mean, +1SD
  Johnson-Neyman output

```
DATA LIST FREE/
  MEAN_PRE   Perceive   Difficul   MEAN_POS   .
BEGIN DATA.
   -.9323    -.7420   -3.9123    4.0172
    .0000    -.7420   -3.9123    5.1559
    .9323    -.7420   -3.9123    6.2946
   -.9323    -.7420     .0000    4.1712
    .0000    -.7420     .0000    5.1230
    .9323    -.7420     .0000    6.0749
   -.9323    -.7420    3.9123    4.3251
    .0000    -.7420    3.9123    5.0902
    .9323    -.7420    3.9123    5.8552
   -.9323     .0000   -3.9123    4.3384
    .0000     .0000   -3.9123    5.2994
    .9323     .0000   -3.9123    6.2604
   -.9323     .0000     .0000    4.4923
    .0000     .0000     .0000    5.2665
    .9323     .0000     .0000    6.0408
   -.9323     .0000    3.9123    4.6463
    .0000     .0000    3.9123    5.2337
    .9323     .0000    3.9123    5.8211
   -.9323     .7420   -3.9123    4.6595
    .0000     .7420   -3.9123    5.4429
    .9323     .7420   -3.9123    6.2263
   -.9323     .7420     .0000    4.8135
    .0000     .7420     .0000    5.4100
    .9323     .7420     .0000    6.0066
   -.9323     .7420    3.9123    4.9674
    .0000     .7420    3.9123    5.3772
    .9323     .7420    3.9123    5.7870
```

```
END DATA.
GRAPH/SCATTERPLOT=
 MEAN_PRE WITH     MEAN_POS BY       Mean_suc /PANEL   ROWVAR=
Mean_dif
```