Methodological Transfer of the "Extended Validation Protocol" in a Human Factors and Usability Engineering Approach to Decision Support Systems in the Sales Domain.

Master Thesis (25EC)

Saskia Henrichs
University of Twente

Department of Cognitive Psychology and Ergonomics

Date: 24th September 2019

Supervisors: Dr. Simone Borsci
Dr. Martin Schmettow

Abstract

Making decisions in increasingly complex environments is a challenge for decision-makers due to cognitive information-processing limitations. The sales domain is exemplary for situations in which decisions need to be made under high time-pressure and high pressure to succeed. With increasing competition for marketing dominance, many companies have implemented Decision Support Systems (DSS) in the field of sales (Roh, Ahn & Han, 2005). Despite the increasing popularity, many DSSs have failed so far due to poor design and usability (Horsky et al., 2012). It is therefore crucial for a successful implementation of DSS to have a usability validation framework. However, research on usability validation frameworks in the sales domain is limited and existing ones show methodological flaws (Mysiak, Giupponi & Rosato, 2005; Schmettow, Schnittker & Schraagen, 2017). The rationale of this study was to approach this gap. Schmettow, Schnittker and Schraagen (2017) successfully proposed an extended validation protocol for evaluating medical DSSs addressing current methodological shortcomings. The goal of the study was to assess a methodological transfer of the extended validation protocol into the sales domain. For this purpose, a case study has been carried out. During a conceptualization phase, a novel design has been found with the help of expert interviews. In the usability testing phase, the novel design has been validated by using the old design as a reference measurement. Statistical analysis of usability tests followed the reference regression model proposed by Schmettow, Schnittker and Schraagen (2017).

Qualitative analysis revealed, that with the novel design, still a number of usability problems remained. Nevertheless, quantitative comparison of the designs showed an overall improvement in efficiency of use and learnability for the novel design, which suggests that a methodological transfer of the framework by Schmettow, Schnittker and Schraagen (2017) into another domain is potentially possible.

1. Introduction

Decision Support Systems (DSS) provide a promising solution to improving decision making. DSSs are interactive systems helping decision-makers to solve problems through utilizing data and models (Ford, 1985). These systems provide support in semi-structured and unstructured situations by bringing together computerized information and human judgement (Asemi, Safari & Zavareh, 2011).

The principal objective of DSSs is to support human decision-makers during the process of arriving at a decision (Uma, 2009). Making increasingly better and more informed decisions is inevitable to be successful in today's competitive society. Adopting discovery knowledge techniques in order to work out these decisions is most often impractical as they require an incredible amount of expertise (Zorilla & García-Saiz, 2013). Furthermore, decision environments are habitually complex. This complexity originates mainly in a large number of parameters which influence a decision, numerous or loosely defined options and high time constraints imposed upon a decision-maker (Bohanec & Rajkovič, 1990). Information-processing limitations are therefore likely to have a deleterious impact on performance. Due to the complexity of decision-making environments, decision makers are prone to make use of mental effort reducing heuristics such as anchoring or adjustment. Adapting to the situation by using heuristics can be economically effective (Van Bruggen, Smidts & Wierenga, 1998). Following the "Law of Fluency":

"'Well'-adapted work occurs with a facility that belies the difficulty of the demands resolved and the dilemmas balanced" (Woods & Hollnagel, 2006).

Hence, making use of mental effort reducing strategies can also lead to systematic and predictable errors (Van Bruggen, Smidts & Wierenga, 1998). Therefore, raising information needs in increasing market dynamics and the potential of new technologies with high-capacity storage media, require efficient information management in the form of DSSs (Wöber, 2003).

As decision making problems occur on a daily basis, in almost any field of human activity, there is a broad scope of potential application domains for DSSs (Bohanec & Rajkovič, 1990; Kersten, Mikolajuk & Yeh, 2000). Kersten, Mikolajuk and Yeh (2000) identified several application fields of DSSs. To name some of them, these domains concern medicine, business and organizational support, environmental decision making and infrastructure. Additionally, a field in which DSSs are extensively used is the domain of sales. Sales agents are mainly using DSSs to drive their conversation with customers or to support their relationship (customer relationship management DSSs). Due to immense competition in this domain, time pressure makes acting fast nearly as important as acting right (Van

Bruggen, Smidts, Wierenga, 1998). Sales DSSs are therefore of particular importance in providing decision-models and forecasts in order to relief the decision-making process in this domain (Wöber, 2003).

This assumption is supported by the study by van Bruggen, Smidts and Wierenga (1998). The authors examined the influence of time pressure and the usage of a sales related DSS in a 2x2 factorial design study. It has been assumed that systematically using all information available with the help of a DSS can compensate for the lack of analytical capabilities of the decision maker and therefore prevent from using potentially erroneous heuristics (Van Bruggen, Smidts & Wierenga, 1998). It was found that pressure indeed had a negative impact on performance through increased complexity of the decision-making environment and therefore lead to superficial information processing. Participants working under time pressure profited most of using a DSS. Furthermore, DSSs were indeed able to compensate for low-analytical capabilities (Van Bruggen, Smidts & Wierenga, 1998). The potential contribution of DSSs in the sales domain is further supported by the study of Achabal, Mclntyre, Smith and Kalyanam (2000). The authors implemented a vendor managed inventory DSS, which improved customer service levels and inventory turnover considerably.

Hence, DSSs are potentially very powerful when well-founded, unambiguous and actionable advice needs to be provided (Horsky, Schiff, Johnston, Mercincavage, Bell & Middleton, 2012). With increasing competition for market dominance, many companies have utilized DSSs in the sales domain, mostly to manage the customer relationship (Roh, Ahn & Han, 2005). Despite their increasing popularity, research has shown that the success of DSS development is uncertain. Although currently a high number of DSSs is being developed, the risk of these systems failing when facing real-world problems is high (Mysiak, Giupponi & Rosato, 2005). So far, many DSSs have failed when facing complex and unstructured problems (Mysiak, Giupponi & Rosato, 2005). Apparently, the performance level of DSSs can be significantly reduced by poor and outdated design. Inadequate design of DSSs may lead to disruptive guidance or even irritating and impeding the cognitive workload of users (Horsky et al., 2012). Concluding, Horsky et al. (2012) state that poor usability is one of the essential impediments to the adoption of DSS and even deterrent to its routine usage.

Consequently, evaluating the usability of DSSs is crucial to the decision-making success and to the continued use of the system. Without proper validation, DSSs can cause costly errors (Borenstein, 1998). However, to our knowledge, research on usability of DSS in the context of the sales domain is limited and evaluation frameworks to assess the usability of these type of systems are missing (Roh, Ahn & Han, 2005). In the healthcare domain however, a framework proposed by Schmettow, Schnittker and Schraagen (2017) revealed promising results. Therefore, the rationale of the present study is to identify an evaluation framework which has already been adopted in another field and to attempt a methodological transfer of the validation protocol by Schmettow, Schnittker and Schraagen (2017) to the field of sales. The procedure will be demonstrated with the help of a case study.

1.1 Issues in DSS interaction due to cognitive limitations

Decision-making environments are usually very complex. Decision-makers are faced with large amounts of complex information which make information-processing limitations likely to influence the decision process. DSSs propose a possibility to compensate for these limitations by managing information more efficiently and handling the accuracy-cost trade-off. Despite potentially being powerful information tools, inadequate design of DSSs may lead to impediments in cognitive workload of a user (Horsky et al., 2012). Understanding the origin underlying this impediment is relevant in order to be able to establish a valid evaluation framework for these systems.

A major challenge that remains with the implementation of DSSs is not to put additional cognitive load on the decision-maker due to the way the information is presented. This challenge is well-known from the literature. Following Woods and Hollnagel (2006), one crucial difference between the clumsy and skilful use new technologies is the disparity between undermining and supporting the processes of workload management. Sweller (2011) described this balancing-act due to human cognitive limitations in more detail in his "Cognitive Load Theory" (CLT). CLT uses evolutionary explanations to consider the human cognitive architecture and to devise novel, instructional procedures based on that architecture.

Essentially, there are two types of instructional cognitive load: intrinsic and extraneous cognitive load. Intrinsic cognitive load is determined by its intrinsic nature of the information while extraneous cognitive load is determined by instructional design factors. Concluding, intrinsic cognitive load executes heavy cognitive load because of its inherent nature (Sweller, 2011). Consequently, this kind of load can only be changed by changing the knowledge level of a person or the nature of the information. The other possibility is, that heavy cognitive load is put on the working memory due to the way the information is presented. This is referred to as extraneous cognitive load (Sweller, 2011).

Given that a person's cognitive capacity is limited, if one kind of cognitive load is too high, the other one needs to be decreased (Mayer, 2005). Referring to the problem at hand, the environment of decision-makers is intrinsically difficult due to a high amount of information and the need for integrating different kinds of sources. Intrinsic cognitive load is therefore unavoidably high. Consequently, only very little cognitive capacity is left for extraneous cognitive load. Information processing and DSS interaction becomes therefore very hard when usability of the system is poor as cognitive capacity needs to be focused on the interaction. When considering the theoretical background of cognitive capacity, it appears logical, that poor usability has been identified as one of the core barriers to the usage of DSSs (Horsky et al., 2012). When providing a validation framework for DSSs, it needs to be taken into account, that due to limited cognitive capacities, a good usability, especially in terms of efficient interaction, is essential for preventing users from being cognitively overloaded.

1.2. Usability issues and adoption barriers to DSSs

Research on human factors and usability engineering has shown that technologies need to be carefully designed if its benefits are to be realized (Karsh, 2004). In line with that, poor usability has been identified to be one of the core barriers to the adoption of DSSs due to its unfavourable effect on extraneous cognitive load (Horsky et al., 2012). Additionally, a study by Tsopra, Jais, Venot and Duclos (2013) showed that an interface designed according to a usability engineering approach was perceived to be more usable and increased confidence of users. A proficient usability is therefore a key requirement for the successful implementation of DSS, independently from the domain and gaining a deeper understanding of the concept is a prerequisite to approach a usability validation framework.

During the last decades, several attributes and criteria for usability emerged (Jeng, 2005). One of the most important benchmarks is provided by the international organisation for standardisation (ISO, 1998). Following ISO 9241-11, usability is defined as:

“The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction, in a specific context of use”.

Effectiveness concerns the accuracy and completeness with which users achieve specified goals. Efficiency on the other hand regards the resources expended in relation to the accuracy and completeness with which users achieve their goals. Satisfaction of use is about the freedom from discomfort and a general positive attitude to the use of the product (Jokela, Iivari, Matero & Karukka, 2003). The ‘context of use’ mentioned in the definition relates to the characteristics of the users, the tasks and the organizational and physical environments of the product (Jokela, Iivari, Matero & Karukka, 2003).

One parameter which is not explicitly mentioned in the ISO 9241-11 definition is learnability. However, learnability is one of the most fundamental usability attributes (Jeng, 2005). It is defined as the ease of use with which features required for achieving goals can be executed. Learnability specifically concerns the capability of a product to enable users to feel that they can productively use the product right away and then quickly learn other new functionalities. Learnability can for example be rationalized as the number of states and state transitions necessary to carry out user tasks (Seffah, Donyaee, Kline & Padda, 2006). For novice users, learnability is especially relevant. In general, a system should be easy to learn in order to enable users to productively work with the system as early as possible (Jeng, 2005).

Concluding, validating usability is a fundamental criterium to the successful implementation of DSSs. The ISO 9241-11 definition can be used as a reference to assess usability. It is generally recommended to seize all of the concepts (effectiveness, efficiency, satisfaction and learnability) to validate the usability of a product. However, the actual choice of criteria strongly depends on the research question at hand (Schmetow, Schnittker & Schraagen, 2017).

1.3 Main deficiencies in DSS assessment methodology

Proper usability validation is important to the decision-making success and to the continued use of DSSs. Without suitable validation, DSSs can lead to costly errors (Borenstein, 1998). Nevertheless, research on DSS evaluation frameworks is still limited in general, especially with reference to the sales domain (Mysiak, Giupponi & Rosato, 2005). Further, usability evaluation methodologies adopted currently have been found to have major flaws (Schmettow, Schnittker & Schraagen, 2017). A first shortcoming concerns indicators for erroneous actions. Deviating from the normative path when interacting with a system is likely to cause additional cognitive workload in situations of time constraints and interruptions. Therefore, the frequency of a user deviating from a normative path can be considered as an indicator for a system's safety. However, even though path deviations and lostness are a widely accepted measure for usability testing, this aspect is seldomly investigated (Schmettow, Schnittker & Schraagen, 2017). A second shortcoming is the assessment of learning. Concurrent studies apparently make conclusions mainly based on single encounter usage of devices. This is not fair, as users are often habituated to an old design. This might lead to a negative transfer and consequently, the new interface performing weaker at a first encounter. Only longitudinal within-subjects studies can track individual differences and learnability (Schmettow, Schnittker & Schraagen, 2017).

In order to challenge these shortcomings, Schmettow, Schnittker and Schraagen (2017) proposed an extended validation protocol in which two extensions have been added to traditional validation testing. The first extension concerns tracing the process of a user's interaction with the system. It entails a novel method to represent a user's task completion process in a more fine-grained way than traditional methodologies (Schmettow, Schnittker &

Schraagen, 2017). Essentially, a task modelling technique from cognitive systems engineering and an algorithm which yields a distance metric for path deviations is applicated (Schmettow, Schnittker & Schraagen, 2017).

The second extension in the validation protocol addresses the methodological flaw of single-encounter studies. Instead, Schmettow, Schnittker and Schraagen (2017) propose a longitudinal research design in which participants conduct multiple sessions. It is accentuated that the same measures need to be applied on variants of the same tasks in order to assess learnability and training effects. For the purpose of evaluating the extent to which users acquired a mental model of the device instead of simply recalling a particular sequence of operations or triggering simple motor-sequence learning, it is advised to create as many task variants as there are sessions (Schmettow, Schnittker & Schraagen, 2017). Additionally, to employ the diversity of individual responses for user-level analysis, it is also strongly proposed to employ a within-subjects design. With the help of this design, individual diversity in training progresses can be captured (Schmettow, Schnittker & Schraagen, 2017).

1.4 Transferring the extended validation protocol to the sales domain

To our knowledge, little research exists on the topic of usability validation frameworks concerning DSSs in the sales domain. Additionally, current usability validation frameworks demonstrated methodological flaws (Schmettow, Schnittker & Schraagen, 2017). Schmettow, Schnittker and Schraagen (2017) addressed these flaws and extended the usability validation protocol for medical devices with promising results. The validation framework by Schmettow, Schnittker and Schraagen (2017) which has already successfully been adopted in the field of healthcare will therefore be adapted to attempt a methodological transfer to the sales domain.

As indicated by the authors, the choice of measurement criteria for usability remain dependent on the research question at hand, which is most likely to be influenced by the application domain (Schmettow, Schnittker & Schraagen, 2017). Therefore, the adopted usability measurements by Schmettow, Schnittker and Schraagen (2017) have been examined for suitability to the domain of interest of the current study.

Following the extended validation protocol, learnability can only be assessed with the help of a longitudinal within-subjects design. As learnability has been assessed as one of the most important aspects of usability, it is only reasonable to follow the recommendations in the extended validation protocol (Jeng, 2005). Furthermore, capturing performance beyond the first encounter with a new system by exploring the impact of task repetition on performance as well as data analysis on a user level is a rational choice, independent of the domain of application.

Besides from learnability, the extended validation protocol proposes deviations from the normative path, cognitive workload and completion times as usability criteria (Schmettow, Schnittker & Schraagen). Satisfaction, as indicated by the ISO 9241-11 definition is not estimated in the reference protocol.

The sales domain is characterised by high time-pressure. An employee using a DSS in this working environment is frequently in real-time situations. It is therefore a logical consequence to take over completion times as a usability criterium for the sales domain. Another factor which is influencing users' interacting with DSS in complex decision-environments are the information-processing and working memory limitations. It has already been analysed, that a system needs to be designed in a way that extraneous workload is enhanced instead of putting additional workload on the user. This applies to the sales domain to the same degree as the healthcare domain. Therefore, cognitive workload is also a relevant usability parameter in validating a sales specific DSS.

However, when concerning measuring effectiveness with the help of the process tracing technique proposed by Schmettow, Schnittker and Schraagen (2017), an adaptation is proposed for the sales domain. Opposing to the healthcare domain, in which operating with a system in the exact normative way is a matter of patient safety, in the sales domain it might not be too critical if the optimal path has strictly been followed to achieve a goal. (Schmettow, Schnittker & Schraagen, 2017). As deviations from the normative path are not assessed as the most crucial performance measurement, the current study will not regard the process tracing technique for pragmatic reasons.

Besides from the adaptation of usability criteria, the proposed procedure in the protocol will also be extended by putting emphasis on the realism of scenarios used during usability validation. Previous research attempted to bring the context of use into a stronger focus by involving testing a system with representative users and tasks in representative settings (Kushniruk, Nohr, Jensen, & Borycki, 2013). Results showed that methods in which realistic scenarios were used under realistic conditions and settings lead to problem-identification which may not have been detected with the help of traditional methods. The evidence from evaluation involving realistic scenarios has been shown to improve the usability and safety of a system (Kushniruk, Nohr, Jensen, & Borycki, 2013). However, due to privacy and ethical concerns as well as the possibility of fatal consequences of errors, in-situ testing is rarely practicable (Svanæs, Alsos, & Dahl, 2010). Consequently, to be able to measure usability beyond traditional desktop user interface evaluation, usability tests need to be conducted in settings that simulate the conditions at a high level of realism (Svanæs, Alsos, & Dahl, 2010). In order to get valid results from such usability tests, use-scenarios need to be realistic. This often requires tests to be run as realistic role-plays with multiple users at the same time (Svanæs, Alsos, & Dahl, 2010).

Additionally, sales agents are frequently in customer contact. It is therefore likely that a DSS in the sales domain will be used as a support tool during the interaction with another person. In these situations, using the DSS is a secondary task, which is mentally demanding. Testing a sales specific DSS in an unrealistic scenario, in which full cognitive capacity can be focused on the tool might not be valid and lead to confounded results.

1.5 Aim of the study

Concluding, DSSs are potentially powerful tools to support decision-makers in complex and fast-paced environments. One application domain in which DSSs are frequently used and are possibly very supportive is the field of sales. Notwithstanding the number of DSSs being developed is increasing, there is a high risk of failing for these systems. The core barrier for the adoption and continued use of a DSS is poor usability (Horsky et al., 2012). Therefore, usability validation is a key success factor in the implementation. Nevertheless, despite its relevance in this field, research on usability validation concerning DSSs is very limited in the domain of sales. Additionally, current usability validation frameworks in general show major methodological flaws (Schmettow, Schnittker & Schraagen, 2017). Schmettow, Schnittker and Schraagen (2017) proposed an extended validation protocol, improving upon these flaws. The study focusses on the usability evaluation of a novel interface of an infusion pump and revealed promising results. The rationale of this study is to identify an evaluation framework which has already been adopted in another field and to attempt a methodological transfer of the validation protocol by Schmettow, Schnittker and Schraagen (2017) to the field of sales.

For this purpose, usability criteria proposed in the extended validation protocol will be adapted with reference to the sales domain. A case study will be carried out in collaboration with a company in which a novel interface of a sales specific DSS will be validated which has recently been implemented by the company (see Appendix A for further information on the case study). This study will be exploratory and therefore investigate if the methodology used in the extended validation protocol can be successfully transferred into another domain by validating an increased usability of a novel interface of a sales specific DSS.

2. Prototyping phase: finding a novel concept

The present study will evaluate the usability of a novel interface of a sales specific DSS. In this phase, the novel version of the interface needed to be conceptualized. Interviews have been conducted with users of the old interface to infer user requirements. User requirements served as input for a prototype of a new interface.

2.1 Methods

2.1.1 Participants

There were three people participating in the interviews. All three participants were male. Participants of the interviews were sales agents of the same company. They were recruited via purposive sampling and data log of activities with the current DSS system of the company was used to get information about their expertise in sales.

2.1.2 Materials. As participants and interviewer did not work in the same location, materials used to conduct the interviews involved the conference tool of the company with a hands-free kit and a screensharing tool in order to bridge the geographical distance. Furthermore, a mobile phone has been used as a recording device. A printout and a digital version of the structured interview-scheme (Appendix B) has been prepared in order to take additional notes by the interviewer. The interview-scheme consisted of 35 questions and each was shown question by question to the participants via a screen-shared presentation. The questions covered seven different sub-topics: the participants assessment of the tool in general, their estimation of the database, questions concerning the portrayal of information, questions concerning the practical value and realism of the tool, special features of the tool and the search function.

2.1.3 Procedure. The interview study has been ethically approved by the ethics committee of the faculty of behavioural, management and social sciences (BMS) of the University of Twente (Request Number 190459).

Instructions and procedure of the interviews were the same for every participant. As participants work in a different location than the interviewer, interviews were conducted via the company-own telecommunication tool. In addition to the interviewer, an observer was present who took notes. Participants were able to see the interviewers via a webcam. Questions were shared via a screensharing program in order to make sure that questions were not misunderstood due to a bad connection. The interviews took place in a pre-booked room at the company, so there were no major distractions. Participants were asked for how long they have been using the tool in order to estimate their level of expertise.

Before the start of the actual interview, it has been emphasized, that the purpose of the interview was not to test the participants' performance, but to test the quality of the tool and that interviewers were only interested in their opinion. Also, participants were made aware of the fact, that their input will be used in the context of the current study. It has been emphasized that all data will be made anonymous. Participants were asked for permission to record the interview and to take notes on the participants answers. Consent was given verbally. The whole interview procedure took about 45 minutes per participant

2.1.4 Data Analysis. Experience of participants with the previous version of the DSS was measured in terms of the number of months they have been working with the tool. The level of expertise has been estimated based on the ratio between the time participants were employed and the average amount of sold deals.

Two of three interviews were recorded and transcribed. The first interview was not recorded as no consent has been given by the participant. However, the participant gave permission to take notes during the interview. Transcripts and notes of the interviews were imported to the program "Atlas.ti". Inductive coding has been used to develop a code-scheme. Interviews were attentively read, and important aspects have been marked by the researcher. Quotes which have been highlighted as important have been sorted per topic. Codes were created based on the lowest common denominator covering all of the quotes assigned to a topic. So, quotes describing the same overarching concept have been assigned to the same code. Also, codes based on the frequency of reporting. A code has only been established with an occurrence of at least 10 times during all interviews. Codes mentioned less frequently have not reflected the lowest common denominator. The coding scheme has been assessed as sufficient when each quote that has initially been highlighted as important was assigned to a code.

A definition for each code has been developed. Variance in each code has been evaluated by reviewing quotes of each participant for the respective code. From each code, user requirements were derived which served as input for different conceptualizations of a novel interface design.

2.2 Results

2.2.1. Participants expertise. The first participant started one month before the interview and just finished his training period. Within the training, he used the tool on a daily basis. For this reason, his level of experience has been assessed as being medium to low. As the first participant started working just one month ago, his level of expertise is low. The second participant has been working at the company for five years, but only started using the tool three months ago. From the log data it can be seen that his activity was medium (1022 logged events). Also, he had to adapt his original way of working when the tool was introduced a few months ago. Therefore, his amount of experience with the tool has been assessed as being medium. Also, his performance based on sold deals has been assessed as being medium by his supervisor. The last person who has been interviewed was assessed as having a high degree of experience as he already got his training with the help of the tool and worked with it for about two to three months afterwards. From the log data, it can be seen, that participant one, with 1388 logged events had the most activities with the tool (activities with the tool in the training phase are not considered in the log data). His level of expertise has been assessed as high as he scored most sold deals ever in the company so far. For a summary of the participants expertise see Table 1.

Table 1. Summary of participants of the interviews, subdivided by ID, gender, amount of experience with the tool and the level of experience estimated in working months and sold deals.

Participant ID	Gender	Amount of Experience	Level of Expertise
1	Male	Medium/Low	Low
2	Male	Medium	Medium
3	Male	High	High

2.2.2 Interviews. The goal of the interviews has been to gather input for a new interface conceptualization. The codes in Table 2 were created based on the most important quotes from the interviews. Each code reflects the overarching concept of at least 10 quotes assigned to it. User requirements can be derived from each code. A full list of the requirement specifications can be found in Appendix C.

Table 2. Codes from user interviews with the respective number of times mentioned and a descriptive definition.

Code	Number of Times Mentioned	Definition
Added Value	39	Assessment of usefulness in daily work practice
Navigation	11	Manner by which user interacts with the system due to how components are organized
Layout of Elements	17	Manner in which elements of the user interface are designed
Way of Usage	15	Means by which users integrate the system in their work-environment

One major requirement for the introduction of a DSS in a working environment is to offer them an added value through supporting users in complex situations. Therefore, it was important to assess if the original tool had a practical usefulness for users in their daily work. With 39 occurrences, was the code "added value" by far the most mentioned and apparently a topic of interest. There was significant variance in this code. In general, assessment of usefulness in daily work was rather mixed. One participant found the tool "*by all means very practic*al" (Participant 1) and used the tool so often that he "*internally walk through the steps and arguments of the tool already*" (P1). However, some aspects of the tool have been estimated as "*not practical*" (P2). Mainly this was due to the fact that "*conversations with a customer can change extremely from one second to another*" (P3) and it is therefore "*sometimes impossible to keep the structure, to keep in the corset of the tool. The result is that we do not use [the tool] anymore because it impedes us too much*" (P2). Participant 3 summarized the dilemma as follows:

"It is simply very difficult to read through longer texts and keep the structure, while talking to a customer. This is really hard from time to time. I am not that capable of multi-tasking on that level. But beside this factor, the tool always helps to get back on track with the conversation"

The overarching requirement of the tool is to offer an added value to its users. When having a closer look at this code, there were three specifications of this requirement. First, the tool needs to offer a structured orientation to manage a customer conversation. At the same time, a second specification would be that the structure of the tool also allows for dynamic usage in order to find information quickly. As the conversation might change in an unpredictable manner, it should also be possible to add or delete information in a dynamic way. Information being more dynamic and allowing for jumping back and forth in the system leads to another specification of the requirement: the system should indicate which topics have already been addressed in order to prevent the user from getting lost.

Analysis of the “navigation” code showed, that the requirement concerning an enhanced navigation was closely related to the requirement of adding value to the tool in the sense that increased navigation will produce a more efficient interaction with the system. The code “navigation” concerned the way in which users interact with the system due to the path it provides. It is essential that users know how to navigate through the tool in order to prevent them from getting lost in the system and to make sure that they easily find the information needed. However, in general all of the participants were satisfied with the current way of navigating through the system (P1, P2, P3). There was some criticism concerning specific functions that inhibit the user from a fully efficient use of the product. This concerned mainly, that it should be visible which topics have already been addressed without forcing the user to stay in the structure of the system: “*It is annoying when you have discussed a topic and the system jumps back to the part which you have not yet explicitly marked as being done*” (P1). However, this requirement is already covered with a specification of the “added value” requirement mentioned before. Therefore, no additional requirement will be added for the “navigation” code.

In addition to the codes “added value” and “navigation”, the manner in which elements are designed and integrated in the user interface (UI) of the tool has been assessed an important requirement because poor design could lead to extraneous mental effort. The code “layout of elements” was mentioned 17 times during the interviews. In general, there was consent between participants, that the design was not very clear. The design was assessed as being “*too bulky*” (P1), as elements are “*too big for the eye*” (P1). Participants one and two also found the design “*confusing*”, as the purpose of different design elements could not easily be discriminated. There were even some elements which have not been seen at all by the users: “*actually, I have never been seen this before… This would indeed be very practical to use in a conversation*” (P3). From analysing this code, three specifications of the requirement have been derived. First of all, elements and content in the tool need to be clearly organized and easy to use. Second, elements need to be discriminable per function. Third, to prevent elements from being overseen or appearing too “bulky”, the space available in the interface should be used in an optimal way.

In order to build a usable tool, it is necessary to understand the way in which it is integrated in a working routine. The code “way of usage” occurred 15 times during the interviews and concerned the point in time as well as the purpose of usage. There was noteworthy variance in this code as well. The way in which the tool is being used differed from user to user. The first participant stated to use the tool “*before a call, in terms of preparation. Rarely during a conversation*” (P1). However, it was also mentioned by the same participant, that “*when I get into the office in the morning, logging in is the first thing I am doing*” (P1), which indicates that even though the tool is used as a preparation for a conversation, it nevertheless has a great importance for the user. The second participant mentioned to use the tool “*before a conversation and during a conversation*” (P2). Participant 3 even mentioned another functionality of the tool: “*it does make sense to look at [the tool interface] before the conversation and during the conversation but also in terms of training sessions, to keep updated all of the time. So, I basically use it all of the time*” (P3). It is therefore required, that the tool delivers information for different purposes: usage for preparation, usage during the conversation and usage for training sessions.

2.2.3 Prototypes. Based on the results from the interviews, a list of requirements has been extrapolated (see Appendix C) to drive the development of the prototype. Based on this knowledge, it has been tried to anticipate a design which followed the requirements. Each wireframe differed in the conceptualization of elements. To be able to find the most suitable design, each design was mapped against the list of requirements.

Figure 1 shows the mock-up which has been conceptualized in the prototyping phase. With this mock-up, all of the requirements were met. For this reason, wireframe 1 has been chosen for further elaboration. To get an overview on the wireframes which have been discarded as well as the final prototype, see appendix D.

2.3 Conclusion

Results of the interviews showed, that the DSS was not usable in a daily work context mainly in terms of efficiency of use. The design of the tool has been assessed as ‘bulky’ and ‘unclear’, which might have a negative impact on the efficiency of use. The logical consequence of a tool being not efficient, is a detrimental effect on the perceived usefulness and the way the tool is used (or not used at all).

What has been striking during the interviews was, that participant one and three, who were relatively new in the company and therefore already received their training with the tool, found the tool more practical and used it more often. It appeared that habituation to the tool is relatively complex for people who did not receive explicit training.

Based on the interviews, a list of user requirements has been established. These requirements have been used to drive the development of a prototype. The prototype was used to conduct usability tests comparing the legacy and the novel design. The following section describes the usability testing phase in more detail.

3. Usability testing: comparing legacy and novel design

In the usability testing phase, a comparative analysis of the original tool and the prototype created in the previous phase has been carried out to evaluate the usability of the novel design. In order to establish the tasks for the usability tests, requirements from the interviews have been analysed by the researcher in more depth. For this purpose, the use case, requirement type, the goal and the pre- and post-condition have been assessed per requirement which was to be assessed in the usability test (Appendix E). Based on this information, a respective task has been selected which reflected one requirement each. Also, the solution to this task has been noted down in order to be able to analyse any deviations from this optimal solution in retrospect. As using the tool under scrutiny is, in practice always a secondary task, a valid assessment of usability had to be made in a realistic setting. Therefore, the defined tasks were transformed into a realistic scenario, in which a shortened version of a customer conversation was played. Hence, tasks were not mentioned explicitly to participants, but intertwined in the scenario.

3.1 Methods

3.1.1 Participants. Fifteen participants were involved in the usability test of the prototype. Participants were recruited via a convenience sample, which included fellow students of the researcher, colleagues and sales agents who already participated in the interview study. Participants 11, 12 and 13 from the usability study were the same as participant 1, 2 and 3 from the interview study (in the same order). For a summary of people participating in the usability tests, see Table 2.

Table 2. Summary of participants of the usability tests subdivided by ID, which session they started with and occupation.

Participant ID	Condition	Occupation
1	A	Student
2	A	Student

3	B	Student
4	B	Student
5	A	Student
6	B	Design/UX
7	B	Scrum Master
8	A	Developer
9	B	Design/UX
10	B	Design/UX
11	A	Scrum Master
12	A	Design/UX
13	A	Sales Agent
14	A	Sales Agent
15	B	Sales Agent

3.1.2 Materials. Two printouts of the informed consent were provided to each participant. The informed consent (Appendix F) covered the nature, purpose and procedure of the experiment. Also, anonymity was accentuated in the consent form and the participant was informed that he may stop the experiment at any time without naming reasons. A printout of more detailed information on the scenario the participant will be facing and the role to play was provided to the participant. As an orientation, the researcher had the tasks and bullet points of the scenario written down to be able to lead the scenario. In order conduct the usability tests, a laptop was provided by the researcher to show the legacy tool and the novel interface to participants. The whole procedure was recorded via a screen recording program which recorded the screen of the laptop as well as the voices of the researcher and the participant. In the end, a digital version of the "Rating Scale Mental Effort" (RSME) (Appendix G) was shown to participants on the same laptop were the interfaces were shown. The RSME is a one-dimensional scale which measures mental effort on a continuous vertical line. Nine labels are disposed on the line ranging from zero to 150. The labels range from "absolutely no effort", to "rather much effort" to "extreme effort" (Paxion, Galy & Berthelon, 2014).

3.1.3 Procedure. A 2x3 (design x session) within-subjects design was carried out. Usability tests involved six sessions. Three sessions were done with the old interface and three with the new one. Half of the participants started with the old interface (Condition A) and half of the participants started with the new interface (Condition B). The condition was randomly assigned. Each session involved one complete customer conversation (scenario), followed by a subjective rating of mental effort on the "Rating Scale Mental Effort" (RSME) directly after the scenario.

Usability tests have been ethically approved by the ethics committee of the BMS faculty of the University of Twente (Request Number 190459). The following procedure was tested in a pilot with one participant before the actual usability tests (Appendix H). Usability tests were conducted in a calm environment with no major distractions. Initially, participants received information on the nature of the DSS, the purpose of the study and the exact procedure of the usability test. It has been emphasized that all data was going to be handled anonymously and that participants may stop the experiment at any time without naming any reasons. Participants were informed about the purpose and handling of recordings from usability tests. Two printout versions of the informed consent were given to each participant. After assuring that participants had fully understood everything and after the possibility to ask questions, they were asked to sign the informed consent. A copy of the consent form was given to each participant. Afterwards, specific instructions on the handling of the tool and important information needed for the scenario were provided to the participant. After the last usability test, participants were thanked and made aware of the fact that the researcher's contact data was mentioned on the consent form in case of further questions or remarks.

3.1.4 Data analysis following the reference regression model. Statistical modelling followed the suggestions of the reference regression model by Schmettow, Schnittker and Schraagen (2017). Initially, non-parametric analyses have been carried out to explore the data. For regression analyses a linear mixed-effects model (LMM) has been fitted to the data, as this model suited the repeated-measures design of the study (Schmettow, Schnittker & Schraagen, 2017). The LMM allows to draw conclusions not only about fixed effects but also about the participant variation (random effects). Fixed effects have been calculated to interpret levels of the two proposed designs. However, fixed effects showing an advantage or disadvantage of the novel design would not give sufficient insight on whether this effect equally applies to all participants. Therefore, random effects have been calculated as well to capture the amount of variation in the data. High levels of random effects implicate high dissimilarity in results per participant. Furthermore, the longitudinal test design made analyses about learnability possible. By setting treatment contrasts, the regression model analyses the average performance of the reference design as an intercept at the first session as well as the magnitude of difference to this reference level over the remaining sessions. The output of the analysis is therefore a regression table, indicating an average intercept value with the reference design, the amount of difference with the respective other design and the extent of change towards sessions with both designs. As these regression tables can be inconvenient to interpret, the reference regression model also proposes a variant to create interaction plots to visualize the results in a more intuitive way (Schmettow, Schnittker & Schraagen, 2017).

To demonstrate both variants, following the study by Schmettow, Schnittker and Schraagen (2017), these two different approaches have been followed to analyse mental workload and completion times. For mental workload, fixed and random effects have been calculated by carrying out regression analysis with a gaussian distribution. Actually, Schmettow (2019) proposes beta distribution with a logit link function being appropriate for response variables which are bounded on both sides, such as the RSME scores in this case. However, even if more appropriate models might exist, the reference model used the gaussian model for practical reasons (Schmettow, Schnittker & Schraagen, 2019). For completion times on the other hand, an interaction-only version of the model has been used. Hereby, regression analysis has been carried out and visualized graphically in an interaction plot showing group means, as well as posterior distributions per session.

3.1.5 Data analysis usability problem identification. In order to analyse the data also in a qualitative way, recordings of the usability tests have been evaluated retrospectively. Errors have been determined per participant. For each error, the number of times it occurred, in which session it occurred, and a more detailed description of the error context has been noted. As there was not only the optimal path to get through a scenario, but the system allowed several possible paths, errors have been defined as actions which did not directly lead to the intended location in the interface and were not part of any possible path. Also, it has been noted if participants were lost in the system after the error or if they were able to recover from the error quickly. A participant has been assessed as being lost, when they needed help from the researcher to recover from the error or when they deviated with more than ten clicks from a possible path to reach their goal. All of this data has been summarized in a table for further inspection.

After this initial review of the recordings, a usability problem matrix has been created based on the data collected (Appendix I). Usability matrix were used to estimate the total amount of occurrences of each usability problem and to prioritize problems.

3.2 Results

Quantitative results have been analysed following the reference regression model proposed by Schmettow, Schnittker and Schraagen (2017). Mental workload and completion times were analysed as performance measures. Mental workload has been subjectively assessed by participants with the help of the RSME (scale from 0 to 150). Completion times were measured as seconds to complete one scenario.

3.2.1 Mental workload. Initial data exploration showed a slightly higher mental workload assessments for the new design (89.00, (*95% CI* [63.00; 51.33]) compared to the old one (87.47(*95% CI* [72.00; 67.67]). Figure 2 shows an exploratory plot of individual learning paths. Participants varied remarkably in their scope of assessing mental workload. Following Schmettow, Schnittker and Schraagen (2017) strong variation in random effects can therefore be expected. Visual exploration further shows, that mental workload was indeed essentially the same for both designs. However, participants' training effects over sessions seem to be inconsistent with the old design, while learning curves for the novel design show greater homogeneity. This becomes even more obvious when comparing individual trajectories directly to each other (figure 3). Figure 3 shows clear improvements of the novel design over sessions in terms of training effects for most of the participants.

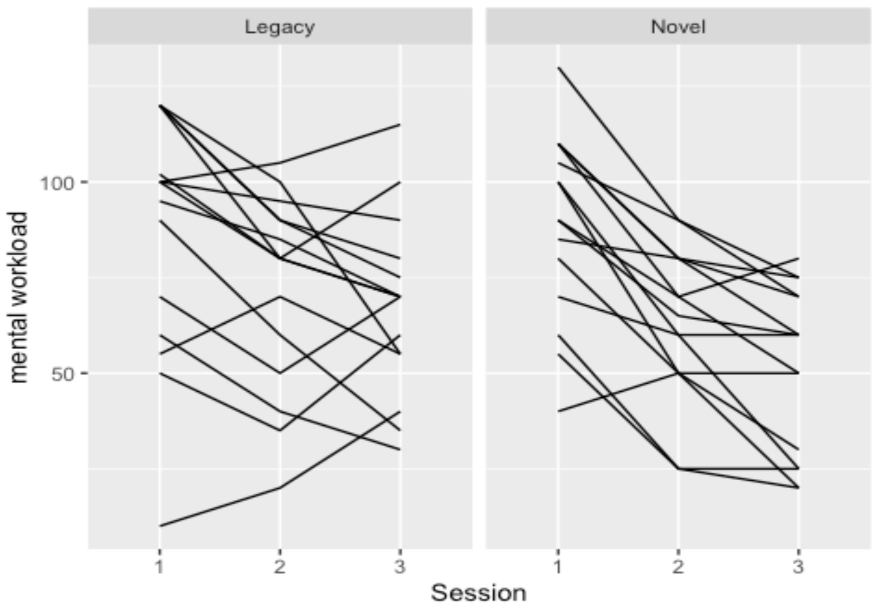


Figure 2. Spaghetti plot showing individual trajectories of mental workload per participant over three sessions.

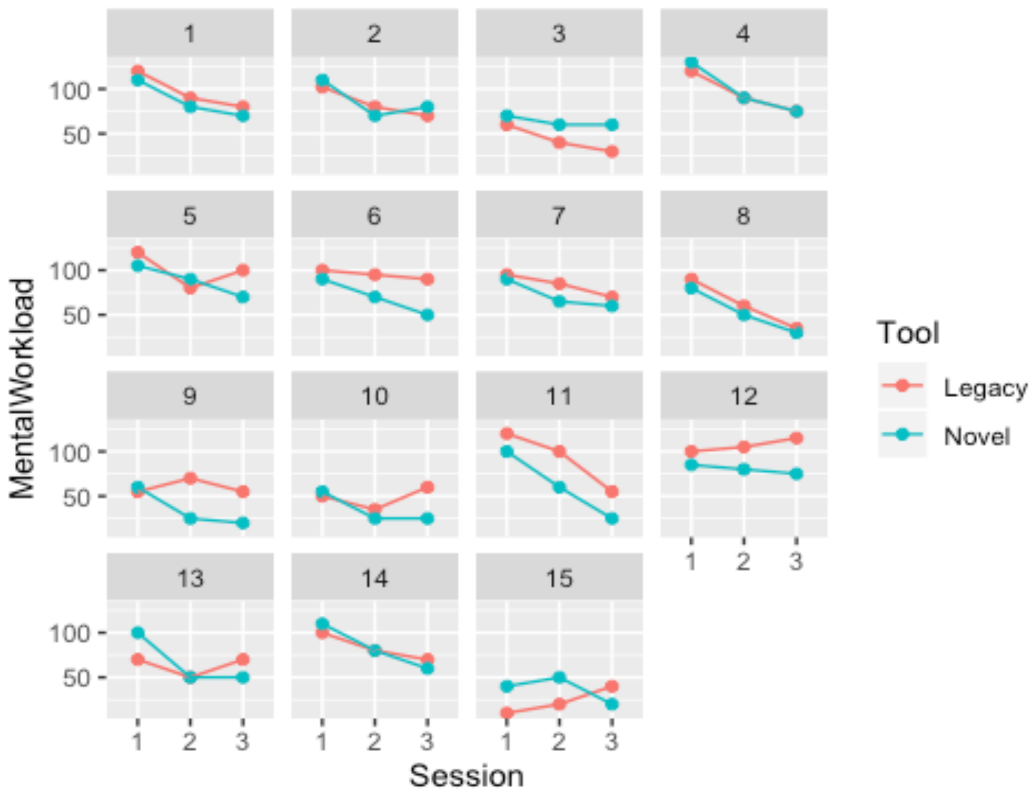


Figure 3. Direct comparison of training effects between legacy and novel design per participant for mental workload.

After initial data exploration, regression results have been calculated. Figure 4 visualizes locations and 95% credibility intervals of fixed effects. The location indicates the central propensity of posterior distributions as well as the area which is most likely to be the true value in terms of group mean or degree of change (Schmettow, Schnittker & Schraagen, 2017). The plot shows only slight differences per design over the three sessions, with a remote disadvantage of the novel design.

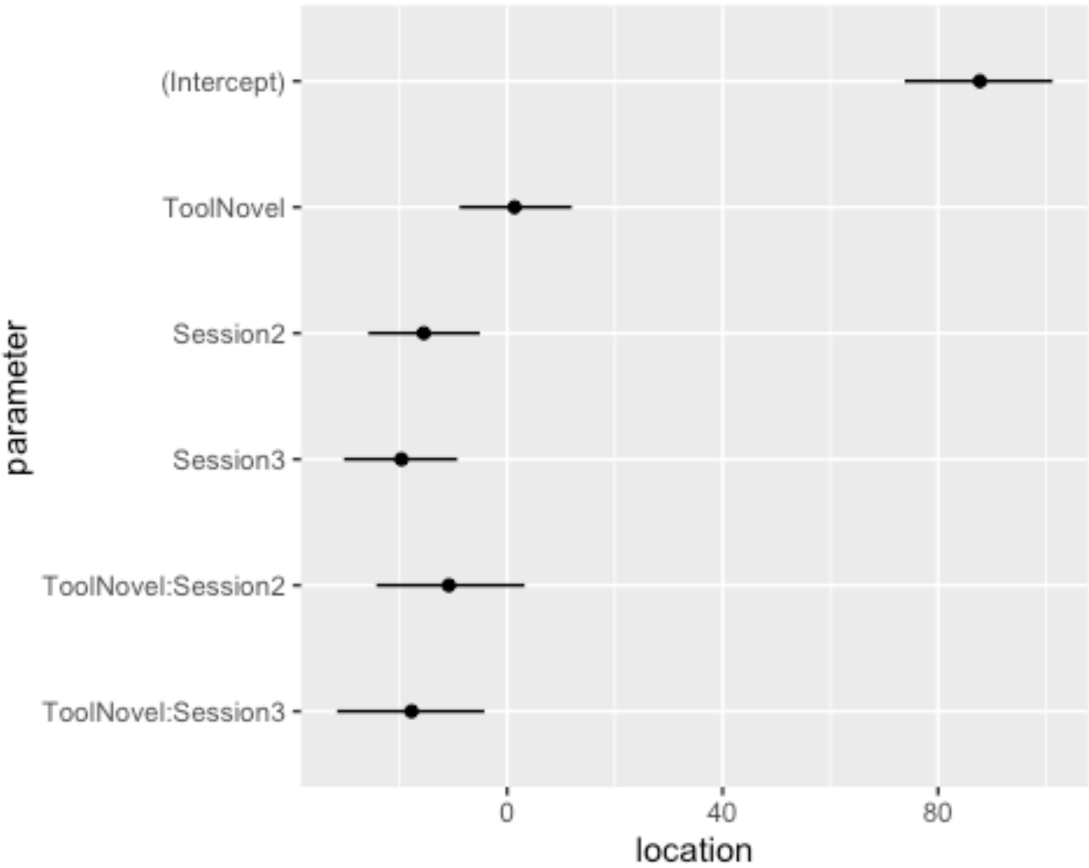


Figure 4. Location and 95% credibility interval of fixed effects for mental workload.

A summary of fixed effects can be found in table 3. At session 1, the novel design revealed slightly higher workload judgements (1.36, *95% CI* [-8.89; 12.01]) than the legacy one. Between session 1 and 2, training happened at a higher rate for the old design (-15.48, *95% CI* [-25.79; -5.09), while improvement rate between session 2 and 3 was in favour of the novel design (-17.76, *95% CI* [-31.59; -4.26]). The comparably steep improvement from session 2 to 3 with the new design in contrast to the smaller effect in the old design might indicate that the old design is approaching the asymptote earlier than the novel design. Therefore, it remains questionable if the novel design might excel the old design with further use. Still, as figure 4 envisions, uncertainty was high for these results.

Table 3. Fixed effects of mental workload with posterior distributions and 95% credibility interval.

Parameter	Location	CI.025	CI.975
Intercept	87.77	73.81	101.23
ToolNovel	1.36	-8.89	12.01
Session2	-15.48	-25.79	-5.09
Session3	-19.63	-30.30	-9.26
ToolNovel:Session2	-10.83	-24.28	3.25
ToolNovel:Session3	-17.76	-31.59	-4.26

As mentioned initially, there was considerable variation in the range in which participants assessed mental workload on the RSME scale. Therefore, random effects analysis has been carried out to examine participant variation. Figure 5 shows the extent of random effect variation as well as residuals. The 95% credibility interval, again indicates uncertainty concerning the estimations.

The figure confirms evident variation between participants in the first session with the novel design (19.50, *95% CI* [12.71; 32.10]). However, participant variation for the old design was basically the same (0.02, *95% CI* [0.00; 13.44]). Concluding, participant variation revealed to be high for both designs. Furthermore, figure 5 shows minor change in variation towards sessions with the novel design. Another aspect which becomes apparent in the plot are high levels of uncertainty concerning these results. Additionally, the plot shows an substantial amount of unexplained variation (13.69, *95% CI* [09.50; 16.66]), which indicates noise in the measurements (Schmettow, Schnittker & Schraagen, 2017). Table 4 shows a summary of quantitative results for random effects.

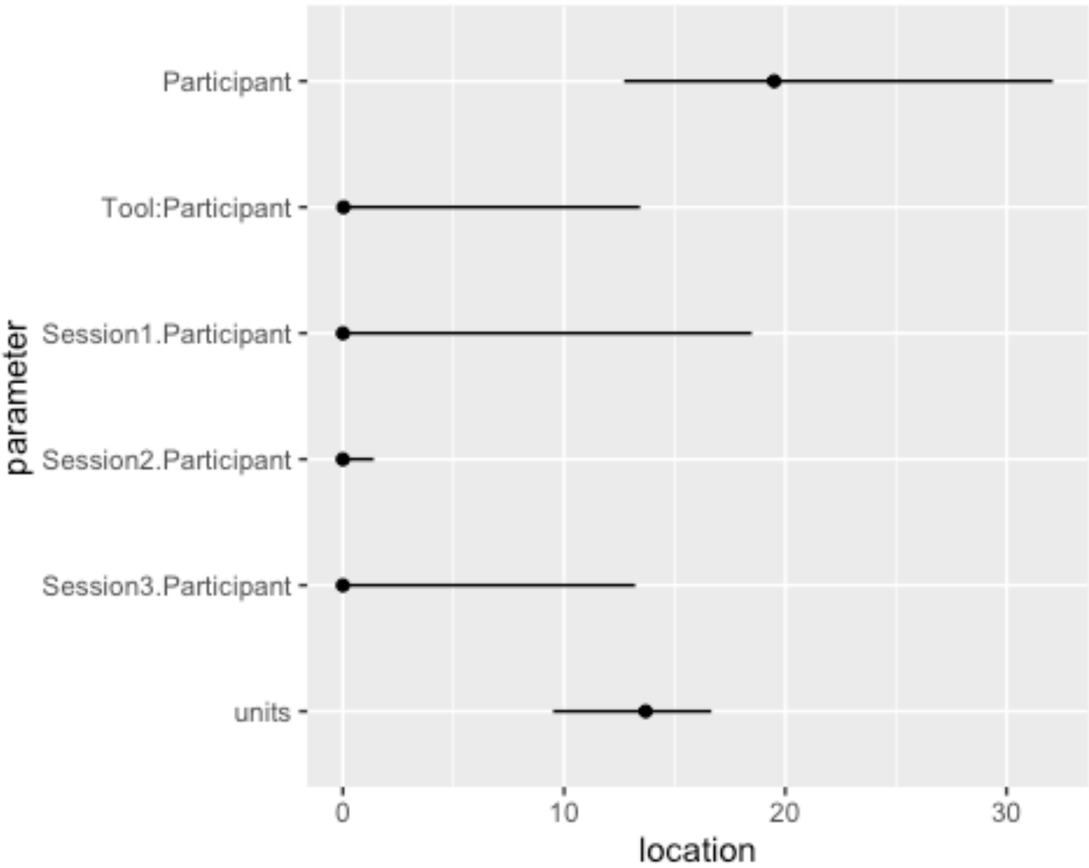


Figure 5. Location and 95% credibility interval of random effects for mental workload.

Table 4. Random effects of mental workload with posterior distributions and 95% credibility interval.

Parameter	Location	CI.025	CI.975
Participant	19.50	12.71	32.10
Tool:Participant	0.03	0.00	13.44
Session1.Participant	0.00	0.00	18.50
Session2.Participant	0.00	0.00	1.40
Session3.Participant	0.00	0.00	13.22
Units (Residuals)	13.69	9.50	16.66

3.2.2 Completion times. Exploratory data analysis indicates that the novel design was superior to the legacy one in terms of completion times in all sessions. The boxplot (figure 6) further shows, that completion times for the novel design had a smaller range than the legacy one in session 2 and 3, indicating less individual variation for the novel design these sessions.

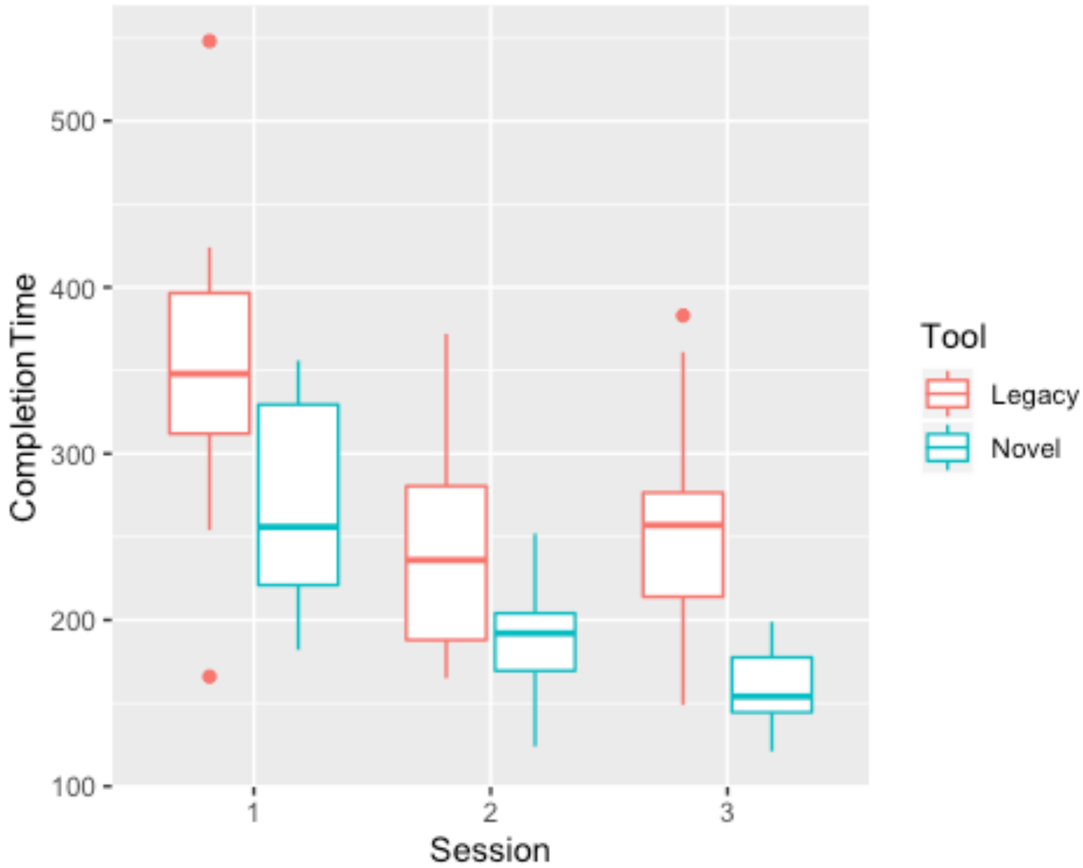


Figure 6. Boxplot showing comparison between novel and legacy design per session.

Visual exploration of individual learning curves supports improvement with the novel design also in terms of learnability (figure 7). Training effects with the novel design show a clearer pattern of improvement from session 1 to 3 than the legacy one.

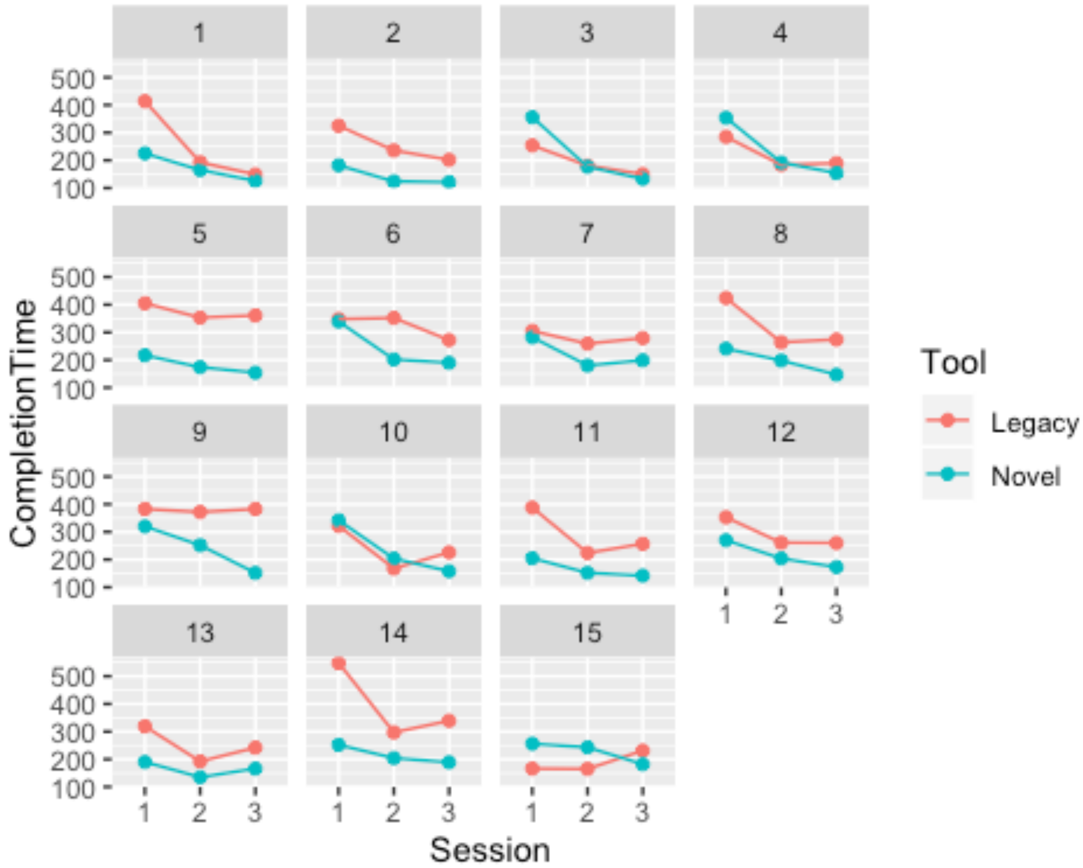


Figure 7. Direct comparison of training effects between legacy and novel design per participant for completion times.

Data analysis for completion times differed somewhat from data analysis for mental workload. For completion times, an interaction-only variant of the reference regression model has been used. Table 5 shows absolute group means of fixed effects for completion times. The interaction plot (figure 8) shows these group means, transformed to the original scale (Schmettow, Schraagen & Schnittker, 2017). The coloured areas in the background are posterior distributions, representing the level of uncertainty. The interaction plot shows the advantage of the novel design in terms of completion times. The novel design seems to improve with continued use, as the asymptote hat not been reached with three sessions. Completion times of the old interface improve even steeper between the first two sessions. At session 2, both interface perform basically equally. However, from session 2 to 3, completion times with the old interface are increasing again. Nevertheless, posterior distributions show considerable overlap and relatively high levels of uncertainty which should be taken into account.

Table 5. Fixed effects for completion times with posterior distributions and 95% credibility interval.

Parameter	Location	CI.025	CI.975
ToolLegacy:Session1	-6.06	-6.38	-5.55
ToolNovel:Session1	-5.76	-6.18	-5.38
ToolLegacy:Session2	-5.66	-6.22	-5.03
ToolNovel:Session2	-5.32	-5.77	-4.92
ToolLegacy:Session3	-5.57	-6.06	-5.07
ToolNovel:Session3	-4.93	-5.53	-4.64

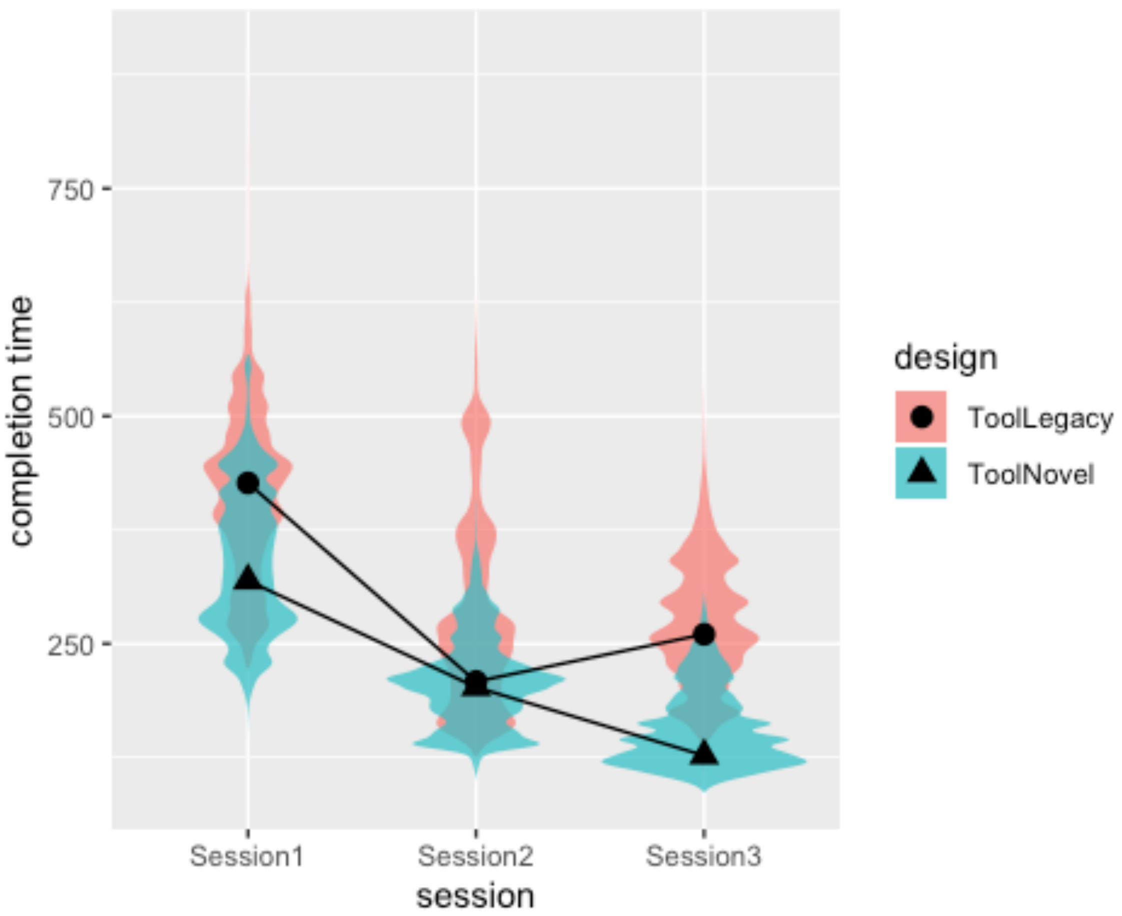


Figure 8. Posterior distributions indicating interaction effects (session x design).

4. Discussion

The rationale of this study has been to approach the gap in usability evaluation research for DSSs in the sales domain, as poor usability has been found to currently prevent DSSs from displaying their full potential in complex decision-making environments (Horsky et al. 2012). The goal hereby was to explore if the methodology proposed by Schmettow, Schnittker and Schraagen (2017) can successfully be transferred from application in the healthcare domain to provide a validation framework for the usability of a sales specific DSS. A new conceptualization of a DSS was based on user research in the form of interviews. In a following step, data analysis following a reference regression model has been carried out in order to compare the novel and the legacy design. Results of the study showed, that the novel interface has been validated in terms of increased usability, which indicates that a methodology proposed by Schmettow, Schraagen and Schnittker (2017) can successfully be transferred into other application domains.

4.1 Interpretation of main findings

Usability measures chosen for interface validation were mental workload and completion times to assess efficiency of use. Concerning mental workload, results were essentially the same for both designs. However, it needs to be kept in mind that the novel interface has been a prototype which was compared to a programmed software. Therefore, there were instances in which the prototype did not work the way participants wanted it to work. When reviewing the recordings, these instances often lead to irritation and rash actions leading to lostness. Lostness and irritation does have a detrimental effect on mental workload (Schmettow, Schnittker & Schraagen, 2017). Concluding, the prototypical nature of the novel interface has led to instances in which participants got lost in the system. It has therefore been unfair to compare a prototype to a programmed software. As mental workload results have been essentially the same for both interfaces, it can be predicted, that mental workload would be less for the novel design, when both interfaces have the same level of fidelity and functionality.

Both designs showed noteworthy participant variation. This might be explainable due to a high spread in responses on the RSME scale. Nevertheless, visualization of individual learning paths has revealed that learning occurred more consistent for the novel design in terms of mental workload. A pattern of steady improvement from session 1 to 3 has been shown homogeneously over participants, while learning with the old design varied widely from participant to participant.

Additionally, analysis revealed, that the novel design showed steeper training effects over sessions compared to the old design. This indicates, that the novel design is likely to bypass the old design with continued use. However, at the same time this means, that the novel interface can still be improved in terms of intuitiveness. Results indicate, that the novel interface is still far from being usable at a first encounter.

Concluding, in terms of mental workload it is assumed that the novel interface will be superior to the old design with comparable degree of fidelity and with continued use. However, results indicated that still, learning is necessary to reach asymptotic use with the novel interface, which implicates that intuitiveness of use can still be improved.

In terms of completion times, the new design exceeded the old one in all three sessions. Concerning results of completion times, the interaction plot showed very irregular posterior distributions. This indicates, that the LMM might not have converged properly. For future analyses, it is therefore proposed to either make use of convergence checks, just as proposed by the authors of the reference regression model or to use an alternative model for completion time analysis. Schmettow (2019) suggests the brm engine with an exgaussian distribution in this case.

Also, from a qualitative point of view, the novel interface performed better than the legacy one. The total amount of usability problems has been reduced from 13 to 10. Additionally, the usability problems found with the novel interface were mainly about aesthetics or minor issues not affecting the basic concepts.

All in all, it can be said that the results in general validated an increased usability. The differences in terms of fidelity and functionality of the compared designs hereby need to be taken into account. However, some uncertainties remain regarding the interaction with continued use, as an asymptote has not been reached with neither design with three sessions. Also, the novel design required some learning effort. Therefore, there is still room for improvement in terms of intuitiveness for the novel design.

The present study showed similar results as the study by Schmettow, Schnittker and Schraagen (2017). Both studies showed that the respective novel design proposed enabled people to perform better and to learn to handle the tool easier. Opposing to the results of the current study, the study by Schmettow, Schnittker and Schraagen (2017) indicated higher workload estimations for the old design in the first session but also higher learning rates. In terms of completion times results of both studies were in line, as the novel interface was superior in both studies. At the same time, results of both studies showed that in general, the novel interface was still far from being usable without any training and therefore leaves room for improvement in terms of intuitiveness.

Concluding, as results of both studies showed a similar tendency of results, it is indicated that the extended validation protocol can successfully be transferred as a usability evaluation framework for the sales domain.

4.2 Strengths and Limitations

Even if the current study basically confirmed that a methodological transfer of the extended validation protocol into another domain is possible, the results were far from perfect. Besides the strengths of the study, there were also flaws which should be considered.

4.2.1 Deviations from the reference model. The current study used the extended validation protocol by Schmettow, Schnittker and Schraagen (2017) as an orientation. However, the current study deviated from the reference model in some respects, which imposes a limitation of this study. A first deviation concerns the proposed process tracing method by Schmettow, Schnittker and Schraagen (2017). The authors used this technique to trace the process of task completion in a user's interaction with a device. Basically, an algorithm is applied which produces a distance metric for path deviations to be able to compare deviations in relation to a normative path. The current study did not make use of this or a comparable accuracy measurement for two reasons. First of all, the technique has not been applied for timely, pragmatic reasons. Secondly, opposing to employees in the healthcare domain, a sales agent's work does not influence patient safety. It is therefore not as important to follow an exact normative path for a sales agent, as long as the goal of selling is reached in the end. Additionally, in reality defining a normative path might not be feasible as sales agents are directly interacting with customers and it is therefore not completely predictable how a conversation will proceed. However, what is indeed very relevant is the effect of lostness when deviating from the optimal path due to its detrimental influence on mental workload (Schmettow, Schnittker & Schraagen, 2017). It is therefore strongly recommended to include either the process tracing method or a comparable technique to identify path deviations in future replications.

Another deviation from the extended validation protocol concerned the analysis of tasks. The reference model included random effects analysis on a task-level in addition to participant-level analysis. However, the current study made use of scenarios for usability testing. Tasks were closely intertwined in the scenarios so it was hardly possible to extract tasks separately for analysis in retrospect. Nevertheless, making use of realistic scenarios was necessary to conduct valid usability testing. As sales agents are frequently in contact with customers, it is likely that they need to use a DSS while interacting with another person and therefore have only limited cognitive capacity to use the system. A usability test which closely resembles a realistic scenario has therefore been indispensable. Nevertheless, it would have been of interest to see the amount difference in task variation between the novel and the legacy interface. Therefore, future research should focus on creating a realistic scenario while at the same time integrating tasks in the scenario which do not overlap and are therefore easily discriminable for retrospective analysis.

4.2.2 Decreasing Levels of Uncertainty. One major limitation of the current study were the levels of uncertainty of the results. One should be careful in making decisive judgements based on results with high levels of uncertainty. For future research, it is recommended to make some effort in improving certainty of results. Increasing the sample size, the number of repetitions and the number of test tasks all increase certainty but at different levels (Schmettow, Schnittker & Schraagen, 2017). A larger sample size will improve the overall certainty of fixed effects. If one is more interested in random effects, as the individual levels of performance, one should increase the number of observations. (Schmettow, Schnittker & Schraagen, 2017). However, once again, the context needs to be taken into account: in high-risk contexts, it is more relevant to get to know the lowest expected level of performance rather than the average level of performance. Reducing the probability of lowest performance will consequently lead to more robust designs. As random effects analysis is able to assess inter-individual variation, it can be used to estimate the robustness of a design (Schmettow, Schnittker & Schraagen, 2017). However, the sales domain cannot be seen as a high-risk context as for example the healthcare domain. From this point of view, one could conclude that increasing certainty of fixed effects with the help of enlargement of the sample size would be sufficient. However, one should note that the current study made use of three sessions, which is the smallest number which can extract non-linearity of training process. Schmettow, Schnittker and Schraagen (2017) propose that if the research goal requires closer determination of training effects, more sessions should be added. For the current study, it would be relevant to know, how many training sessions will be necessary before an asymptotic usage happens. As the asymptote has not been reached with three sessions, it is proposed to add more observations in future research as well. Concluding, when replicating the study, it is proposed to try to improve levels of uncertainty for both random- and fixed-effects by increasing the sample size as well as the number of sessions.

4.2.3 Context of use. One major strength of the current study has been to take the context of use and actual user needs into account when designing a novel interface. As discussed earlier, DSSs do indeed entail the possibility to support humans in overcoming cognitive limitations in complex decision-making environments. However, 'compensating' for information processing limitations is closely related with the risk of the so-called 'substitution myth'. New technological possibilities often cast a shadow over new burdens and complexities which are created for users operating these high-consequence systems (Woods & Hollnagel, 2006). In fact, decision-makers using a DSS act in a sociotechnical system in which humans and technology are closely intertwined in their working environment: a so called joint cognitive system (JCS) (Schmettow, Schnittker & Schraagen, 2017). Following the substitution myth, new technology often simply shifts workload in complex working environments from the human to the technology in order to remove the human from the part of the process and allowing for the workload to drop. The idea that technology can serve as a simple substitution for humans is a serious over-simplification at the blunt end of systems (Woods & Hollnagel, 2006). Often, humans are seen as "end-users" which causes failures in the usability of a design (Shackel, 2009). Instead of attempting to design technology in order to overcome human limits and substitute for human involvement, the critical task is how to design for a resilient and coordinative system that is able to adapt to the demands of the environment and supports humans in their strategies for work (Woods & Hollnagel, 2006).

In order to prevent this, the current study made use of a human factors and usability engineering approach in each validation phase. In the prototyping phase, requirements were based on user interviews. Questions reflected the added value, the contextual use and wishes of participants. During usability tests, realistic scenarios were used which represented the actual working situation of users. The adoption of human factors and usability engineering approach has hopefully led to a successful and resilient integration of the DSS in the working environment of the user without trying to substitute human involvement.

4.2.4 Mixed Research. Purely qualitative studies can be very useful in the identification of design issues. However, the process of observation coding has been assessed as unreliable with immense disagreements between experts (Schmettow, Schnittker & Schraagen, 2017). The research by Mysiak, Giupponi and Rosato (2005) shows, that purely qualitative criteria bring about the risk that users need to explicitly state their requirements and problems. This might lead to overlooking critical aspects in the usability. Quantitative criteria on the other hand such as completion times, mental workload or task completion rates on the other hand can be obtained with little effort and used to validate a system as a whole and with implicit assessment of users' needs and problems (Schmettow, Schnittker &

Schraagen, 2017). However, purely quantitative criteria give little clue on the nature of underlying design issues (Schmettow, Schnittker & Schraagen, 2017).

Approaching the validation of the novel DSS interface with a mixed research was therefore one strength of the current study. The advantages of both qualitative and quantitative research have been used while both methods validated each others' results. Another attempt to bridge the gap between qualitative and quantitative methods was the process tracing technique proposed by Schmettow, Schraagen and Schnittker (2017). In the future, such techniques should preferably be adopted to overcome the limitations of both methods.

4.3 Conclusion and Future Research

The aim of the current study has been to approach a usability validation framework by assessing the feasibility of a methodological transfer of a protocol which has already been adopted in the healthcare domain into the domain of sales. During the course of this study, a human factors and usability engineering approach has been adopted to develop a novel interface for a sales specific DSS. The methodology involved a mixed-research approach in a prototyping and a usability testing phase. During the prototyping phase, requirements analysis and implementation has been done with the help of expert interviews. During the usability testing phase, the novel interface has been compared to the legacy design in a series of usability tests involving realistic scenarios. For this purpose, usability parameters from extended validation protocol by Schmettow, Schnittker and Schraagen (2017) have been adapted to suit the purpose of the current case study. Considering that the interface of the novel design was a prototypical version compared to a fully programmed software, results showed, that the novel interface lead to increased usability in terms of completion time, mental effort and individual patterns of training effects. The results revealed that the extended validation protocol by Schmettow, Schnittker and Schraagen (2017) can successfully be adapted to other domains.

A rather different approach to DSS validation has been adopted in the study by Mysiak, Giupponi and Rosato (2005). The study validated a DSS addressing complex decision problems in the domain of water resource management. The authors made use of a different conceptualization and validation procedure than the present study. Mysiak, Giupponi and Rosato (2005) chose for an evolutionary system development methodology involving the release of three system prototypes. In a series of meetings, expectations of users were collected, and continuous feedback gained. However, in comparison to the current study, users were given an initial prototype from the beginning on, as the authors stated that an initial system to react to would help users to overcome their difficulties in expressing their expectations. The interface then evolved successively according to the suggestions from the first prototype. Essential features have evolved from basic to advanced fidelity towards final implementation (Mysiak, Giupponi & Rosato, 2005). Besides from frequent feedback, the authors do not mention explicit usability testing. In the implementation, ambiguities have been found in user feedback, which resulted in the framework not being readily understandable. The authors therefore proposed a training stage before applying the DSS and using a set of supportive documentation and exercises (Mysiak, Giupponi & Rosato, 2005).

The differences in results compared to the current study show, that a mixed approach in interviewing users for requirements and doing usability testing might lead to more profound validation of a novel interface than purely by explicit user feedback as in the study by Mysiak, Giupponi and Rosato (2005) and is therefore recommended for future research. When presenting users with prototypes and asking for feedback, users might not be able to imagine problems they could be facing, or do not want to offend the researcher with any criticism and therefore confirm what they see. Finding out problems in usability with more implicit testing might lead to more valid results.

As another conclusion, the current study has been about finding a novel concept for the interface. The interface design of the prototype was therefore very minimalistic. Schmettow, Schwabe and Nazareth (2014) showed that visual simplicity and prototypicality increases the experience of fluency of websites, which has a positive effect on the overall usability. It is therefore concluded, that the overall increased usability of the novel interface could have also been influenced by its simplicity. When making further improvements, it should therefore be considered to keep the interface simple when elaborating the interface to a programmed software.

The current study was not without flaws. Further replications need to work on decreasing levels of uncertainty to be able to make conclusive judgements. In general, one can state that the novel interface was more usable than the legacy interface. Nevertheless, there is still room for improvement. Results were still far from the interface being usable without any training. Also, from qualitative error analysis, there were 10 remaining usability problems which deliver input for future improvements of the established concept on a more fine-grained level. Another aspect for further research involves the implementation of the DSS. Expert interviews showed, that the initial user acceptance of the tool was relatively low for some users. Therefore, future research might include the usability parameter satisfaction of use to assess if low satisfaction has increased with the novel interface and if this could have been the reason for low levels of user acceptance. Another possibility would be to focus on user acceptance testing in the future.

References

Achabal, D. D., McIntyre, S. H., Smith, S. A., & Kalyanam, K. (2000). A decision support system for vendor managed inventory. *Journal of retailing*, *76*(4), 430-454.

Asemi, A., Safari, A., & Zavareh, A. A. (2011). The role of management information system (MIS) and Decision support system (DSS) for manager's decision making process. *International Journal of Business and Management*, *6*(7), 164-173.

Bohanec, M., & Rajkovič, V. (1990). DEX: An expert system shell for decision support. *Sistemica*, *1*(1), 145-157.

Borenstein, D. (1998). Towards a practical method to validate decision support systems. *Decision Support Systems*, *23*(3), 227-239.

Ford, F. N. (1985). Decision support systems and expert systems: a comparison. *Information & Management*, *8*(1), 21-26.

Horsky, J., Schiff, G. D., Johnston, D., Mercincavage, L., Bell, D., & Middleton, B. (2012). Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *Journal of biomedical informatics*, *45*(6), 1202-1216.

Iso, W. (1998). 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). *The international organization for standardization*, *45*(9).

Jeng, J. (2005). Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability. *Libri*, *55*(2-3), 96-121.

Jokela, T., Iivari, N., Matero, J., & Karukka, M. (2003, August). The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In *Proceedings of the Latin American conference on Human-computer interaction* (pp. 53-60). ACM.

Karsh, B. T. (2004). Beyond usability: designing effective technology implementation systems to promote patient safety. *BMJ Quality & Safety*, *13*(5), 388-394.

Kersten, G. E., Mikolajuk, Z., & Yeh, A. G. O. (2000). *Decision support systems for sustainable development: a resource book of methods and applications*. Springer Science & Business Media.

Kushniruk, A., Nohr, C., Jensen, S., & Borycki, E. M. (2013). From usability testing to clinical simulations: Bringing context into the design and evaluation of usable and safe health information technologies. *Yearbook of medical informatics*, *22*(01), 78-85.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, *3148*.

Mysiak, J., Giupponi, C., & Rosato, P. (2005). Towards the development of a decision support system for water resource management. *Environmental Modelling & Software*, *20*(2), 203-214.

Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. *Frontiers in psychology*, *5*, 1344.

Roh, T. H., Ahn, C. K., & Han, I. (2005). The priority factor model for customer relationship management system success. *Expert systems with applications*, *28*(4), 641-654.

Schmettow, M., Schnittker, R., & Schraagen, J. M. (2017). An extended protocol for usability validation of medical devices: research design and reference model. *Journal of biomedical informatics*, *69*, 99-114.

Schmettow, Schwabe & Nazareth, D. (2014). *The fluency effect as the underlying variable for judging beauty and usability.* (master's thesis. University of Twente, Enschede, Netherlands.

Schmettow, M., Vos, W., & Schraagen, J. M. (2013). With how many users should you test a medical infusion pump? Sampling strategies for usability tests on high-risk systems. *Journal of biomedical informatics*, *46*(4), 626-641.

Seffah, A., Donyaee, M., Kline, R. B., & Padda, H. K. (2006). Usability measurement and metrics: A consolidated model. *Software quality journal*, *14*(2), 159-178.

Shackel, Brian. "Usability–context, framework, definition, design and evaluation." *Interacting with computers* 21, no. 5-6 (2009): 339-346.

Svanæs, D., Alsos, O. A., & Dahl, Y. (2010). Usability testing of mobile ICT for clinical settings: Methodological and practical challenges. *International journal of medical informatics*, *79*(4), e24-e34.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.

Tsopra, R., Jais, J. P., Venot, A., & Duclos, C. (2013). Comparison of two kinds of interface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics. *Journal of the American Medical Informatics Association*, *21*(e1), e107-e116.

Van Bruggen, G. H., Smidts, A., & Wierenga, B. (1998). Improving decision making by means of a marketing decision support system. *Management Science*, *44*(5), 645-658.

Wöber, K. W. (2003). Information supply in tourism management by marketing decision support systems. *Tourism Management*, *24*(3), 241-255.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. CRC Press.

Zorilla, M., & García-Saiz, D. (2013). A service oriented architecture to provide data mining services for non-expert data minders. *Decision Support System*s, 55, no. 1 (2013): 399-411

Appendix A

Background information case study and old design

This part contains confidential information and is not publicly available.

Appendix B

Interview-Scheme

This part contains confidential information and is not publicly available.

Appendix C

List of requirement specifications

This part contains confidential information and is not publicly available.

Appendix D

This part contains confidential information and is not publicly available.

Appendix E

User test preparation for scenario and task identification based on user requirements

This part contains confidential information and is not publicly available.

Appendix F

Informed Consent

Project Title: **„Methodological Transfer of the "Extended Validation Protocol" in a Human Factors and Usability Engineering Approach to Decision Support Systems in the Sales Domain".**

Investigator: Saskia Henrichs – „Master Student Human Factors and Engineering Psychology"

Supervisor:

Dr. S. Borsci – Cognitive Psychology & Ergonomics

Dr. M. Schmettow – Cognitive Psychology & Ergonomics

Participant Number:...

Participant Name:......................................

Welcome!

First of all, thank you for your interest in participating in this study. With your participation you do not only make a great contribution for finalising a master thesis, but also for validating a product that will actually be implemented and help users in enhancing their job on a daily basis. In the following, you will get an idea on what this study is about. However, please do not hesitate to ask questions before, during or after the study.

Background and Purpose of the Research

Decision Support Systems (DSS) are interactive systems which help decision makers manage data and models (Ford, 1985). There are a lot of domains in which DSS's provide valuable tools to support humans in their daily work. A lot of research has for example been done in the domain of health care. Also, in the sales domain, DSS provide a promising chance to deal with increasing amounts of information in complex decision environments. Approaching the design of DSS from a human factors point of view has proven to be effective (Schmettow, Schnittker & Schraagen, 2017). Schmettow, Schnittker and Schraagen proposed an advanced methodology in order to validating such devices in a safety critical environment. However, by closer looking at it, the model might have its merits for testing other systems. Therefore, in the current study, this model will be subjected to scrutiny in another domain with the help of a case study.

A closer look: the case study

This part contains confidential information and is not publicly available.

What will be your contribution in this study?

Your contribution to this study will be your participation in a usability test. The test consists of two sessions. In each session, we will test one version of the DSS under investigation. Before each session you will get verbal instructions on the procedure of the usability test. Essentially, you will be asked to take over the role of a sales representative and making use of the DSS in a scenario that will be provided by the researcher. The more detailed workwise of the scenario will be explained to you later by the researcher. As the researcher will be part of the scenario, the usability test will be recorded with the camera you see in front of you. The recording will expose your voice. Only the screen of the computer will be filmed, not you interacting with the program. Each scenario will take around 10 minutes. After the usability test, you will be asked to give a short estimation of the mental workload you experienced. The total procedure will take about an hour.

Important Information

Participating in this study will not bring you at risk or discomfort of any kind. Nevertheless, if you feel uncomfortable at any point of the process, you can quit from the study without justification or further consequences. It is hereby stressed, that your participation is completely voluntary.

What is going to happen to your data?

Results will be used in terms of a master thesis and to a corresponding degree public. Results will be presented at the University of Twente and intern in the company. All data that might lead to your identification will be completely anonymised (e.g. your name will be replaced by a number). The recordings of the usability test will be used for retrospective analysis and deleted afterwards. Therefore, no personal data or recordings of your voice will be seen by anyone else than the researcher present. Data will be treated strictly confidential by the researcher.

Contact information

If you want to get access to the results of the study or have any questions left, you can contact the researcher under the following E-Mail address:

Saskia Henrichs
s.henrichs@student.utwente.nl

Agreement

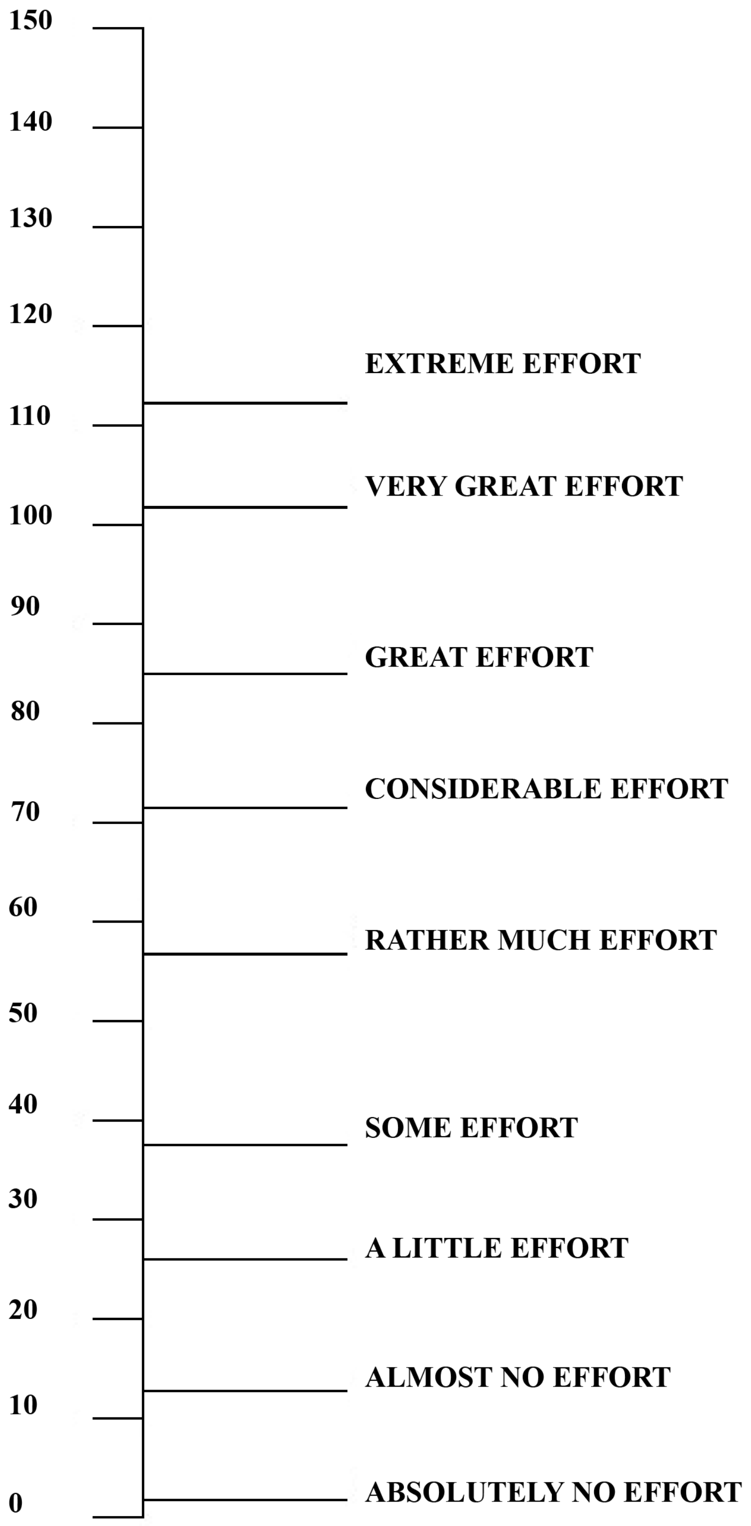
I hereby agree to take part in this study. I have read all the information above and I have been sufficiently provided with all the information I needed to know so far. I declare to have fully understood the content and purpose of the study and what is demanded with my participation. I have been sufficiently informed about the voluntary nature of my participation and the confidentiality of my data. I have received a signed copy of this informed consent (Henrichs, 2018).

Date, Signature Participant

Date, Signature Researcher

Appendix G

Rating Scale Mental Effort (RSME)



Appendix H

Pilot Test

Before the actual data gathering, a pilot test has been conducted in order to test the feasibility of the different scenarios. A fellow colleague (female, 22 years-old) participated in the pilot study. The procedure described for usability tests has been followed for this purpose. However, as the purpose has been to test the feasibility of the scenarios and to estimate the time necessary for the actual tests, no data has been recorded during the pilot test. The researcher took notes on problems occurring during the procedure. Also, the pilot test served as a training for the researcher to become more secure in the role of the customer. The pilot test showed, that the scenarios were realistic and in a practical time-frame. However, there were some points in which the researcher did not manage to follow the scenario and got lost. Critical incidents have been noted and will be attended to during actual usability tests.

Appendix I

Usability Problem Matrices

This part contains confidential information and is not publicly available.

Appendix J

R Markdown Quantitative Data Analyses following the reference regression model proposed by Schmettow, Schnittker & Schraagen (2017). The original markdown of the reference model can be found in the appendix of the article

DataAnalysis

Saskia Henrichs

9/3/2019

```
options(repos=structure(c(CRAN="http://cran.us.r-project.org")))


library(BiocManager)

BiocManager::install("genefilter")

## Bioconductor version 3.9 (BiocManager 1.30.4), R 3.6.1 (2019-07-05)

## Installing package(s) 'genefilter'

##
## The downloaded binary packages are in
##  /var/folders/np/pf_cqd5n0bj64jbq2_t2nqjh0000gn/T//RtmpMpQWbq/downloade
d_packages

## Update old packages: 'tidyr'

library(MCMCglmm)

## Loading required package: Matrix

## Loading required package: coda

## Loading required package: ape

library(modeest)
library(MASS)

library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(tidyr)

##
## Attaching package: 'tidyr'

## The following object is masked from 'package:Matrix':
##
##     expand

library(stringr)

library(ggplot2)
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

##
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':
##
##     nasa

library(knitr)

library(tidyverse)

## Registered S3 method overwritten by 'httr':
##   method         from
##   print.response rmutil

## ── Attaching packages ──────────────────────────────────────────
─ tidyverse 1.2.1 ──

## ✓ tibble  2.1.3     ✓ purrr   0.3.2
## ✓ readr   1.3.1     ✓ forcats 0.4.0

## ── Conflicts ──────────────────────────────────────────── tidy
verse_conflicts() ──
## ✖ tidyr::expand() masks Matrix::expand()
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ✖ dplyr::select() masks MASS::select()

library(devtools)
```

```
## Loading required package: usethis

##
## Attaching package: 'devtools'

## The following object is masked from 'package:BiocManager':
##
##     install

library(mascutils)

##
## Attaching package: 'mascutils'

## The following object is masked from 'package:base':
##
##     mode

library(bayr)
library(readxl)
library(kableExtra)

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows

## MCMC effort (set to lower values for test runs)
nitt = 55000
burnin = nitt/11
```

#Functions

```
#'Logit and inverse logit (link) functions
#'
#'\code{logit} returns the logit and
#' \code{inv.logit} returns the inverse of logit (aka logistic function)
#'
#'@param mu vector of real numbers with largest value \code{upper}
#'@param eta vector of real numbers
#'@upper upper bound of scale (default: 1)
#'@return \code{logit}: real number, \code{inv.logit}: real number [0;uppe
r]
logit = Vectorize(function(mu, upper = 1) log(mu/(upper-mu)))
inv.logit = function(eta, upper = 1) plogis((eta/upper))



#'Subscripts for model parameters
#'
#' Subscripts will be used for fixed effects (BETA) and random effects and
#' residuals (SIGMA) of the reference model
BETA <-
  data_frame(beta = c("0", "D", "S[1]", "S[2]", "S|D[1]", "S|D[2]"),
```

```
             parameter = c("(Intercept)",
                           "designLegacy", "session2","session3",
                           "designLegacy:session2", "designLegacy:session3
"),
             order = c(1,2,3,4,5,6))

## Warning: `data_frame()` is deprecated, use `tibble()`.
## This warning is displayed once per session.

SIGMA <-
  data_frame(sigma = c("P", "P|D",
                       "P|S[1]", "P|S[2]", "P|S[3]",
                       "T", "T|D",
                       "\\epsilon[1]", "\\epsilon[2]", "\\epsilon[3]"),
             parameter = c("Participant", "design:Participant",
                           "1.Participant", "2.Participant", "3.Participan
t",
                           "Task", "design:Task",
                           "1.units", "2.units", "3.units"),
                   order = c(1:10))




#' Summary of fixed effects
#'
#' Fixed effects of the reference regression model are summarized by
#' a location paramater and the 95% credibility interval.
#' The table can be decorated with subscripted parameter names
#' for knitr output
#'
#'@param object MCMCglmm (regression results)
#'@param mean.func mean function to transform the linear predictor (identi
ty)
#'@param loc.func location function for summarizing the posterior (mode)
#'@param format.func function that formats numbers (identity)
#'@param greek vector with subscripts for betas
#'@param order vector giving a different order of betas


fixef <-
  function (object,
           loc.func = shorth,
           mean.func = identity,
           neat.num = F,
           add.greek = F,
           BETA = NULL) {

    if(is.null(BETA)) BETA <-
      data_frame(parameter = colnames(object$Sol))

    if(is.null(BETA$order)) BETA$order <- c(1:nrow(BETA))

    out <-
```

```
      as.data.frame(object$Sol) %>%
      gather("parameter", "value")%>%
      mutate(value = mean.func(value)) %>%
      group_by(parameter) %>%
      summarize(location = loc.func(value),
                CI.025 = quantile(value, .025),
                CI.975 = quantile(value, .975)) %>%
      right_join(BETA, by = "parameter") %>%
      arrange(order) %>%
      mutate(parameter = factor(parameter, levels = parameter))

    if(neat.num) {
      out <-
        out %>%
        mutate(location = neatNum(location),
               CI.025 = neatNum(CI.025),
               CI.975 = neatNum(CI.975))
    }

    if(add.greek & !is.null(out$beta))  {
      out <-
        out %>%
        mutate(beta = paste0("$\\beta_{", beta, "}$")) %>%
        select(beta, parameter, location, CI.025, CI.975)
    } else {
      out <-
        out %>%
        select(parameter, location, CI.025, CI.975)
    }
    out
  }

# fixef(glmm_mentalWorkload, neat.num = T, add.greek = T, BETA = BETA)$par
ameter
# fixef(glmm_pathDeviation, neat.num = T, add.greek = T)

#' Summary of random effects
#'
#' Random effects of the reference regression model are summarized by
#' a location paramater and the 95% credibility interval.
#' The table can be decorated with subscripted parameter names
#' for knitr output
#'
#'@param object MCMCglmm (regression results)
#'@param sd values are given as standard deviations (T)
#'@param loc.func location function for summarizing the posterior (mode)
#'@param format.func function that formats numbers (identity)
#'@param greek vector with subscripts for sigmas
#'@param order vector for reordering the rows

ranef <-
  function(object,
           loc.func = shorth,
           neat.num = F,
```

```
           add.greek = F,
           SIGMA = NULL) {

    if(is.null(SIGMA)) SIGMA <-
        data_frame(parameter = colnames(object$VCV))

    if(is.null(SIGMA$order)) SIGMA$order <- 1:nrow(SIGMA)

    out <-
      as.data.frame(object$VCV) %>%
      as_data_frame() %>%
      gather("parameter", "value") %>%
      mutate(value = sqrt(value)) %>%
      group_by(parameter) %>%
      summarize(location = loc.func(value),
                "CI.025" = quantile(value, .025),
                "CI.975" = quantile(value, .975)) %>%
      right_join(SIGMA, by = "parameter") %>%
      arrange(order) %>%
      mutate(parameter = factor(parameter, levels = parameter))

    if(neat.num) {
      out <-
        out %>%
        mutate(location = neatNum(location),
               CI.025 = neatNum(CI.025),
               CI.975 = neatNum(CI.975))
    }

    if(add.greek & !is.null(out$sigma)) {
      out <-
        out %>%
        mutate(sigma = paste0("$\\sigma_{", sigma, "}$")) %>%
        select(sigma, parameter, location, CI.025, CI.975)
    } else {
      out <-
        out %>%
        select(parameter, location, CI.025, CI.975)
    }
    out
  }

# ranef(glmm_mentalWorkload, neat.num = T, add.greek = T, SIGMA = SIGMA)
# ranef(glmm_pathDeviation, neat.num = T, add.greek = T)

#' Augmented interaction plot
#'
#' Fixed effects are illustrated as design-by-session interaction plot
#' Degree of certainty is shown as underimposed violin plot
#'
#'@param object MCMCglmm (regression results of interaction-effect only mo
del)
#'@param mean.func mean function to transform back to original scale
#'@param loc.func location function for summarizing the posterior
```

```
#'@param ylab label of y axis as shown in plot
regression_plot = function(object, mean.func = identity,
                           loc.func = shorth, ylab = "value"){
  Sol = as.data.frame(object$Sol)[,1:6] %>%
    gather("parameter", "value") %>%
    separate(parameter, c("design", "session")) %>%
    mutate(session = str_replace_all(session, "session", ""),
           design = str_replace_all(design, "design", "")) %>%
    mutate(value = mean.func(value))

  out = Sol %>%
    ggplot(aes(y = value, x = session, fill = design)) +
    geom_violin(color = NA, position = "identity", alpha = .7) +
    stat_summary(aes(group=design),
                 fun.y=shorth, geom="line") +
    stat_summary(aes(group=design, shape = design, fill = design),
                 fun.y=shorth, geom="point", size = 3) +
    ylab(ylab)
  out
}
```

#Load Data

```
library(readxl)
Data <- read_excel("Raw_Data.xlsx") %>%
  mutate(Tool = as.factor(Tool),
         Session = as.factor(Session))
summary(Data)

##   Participant  Condition              Tool    CompletionTime
##  Min.   : 1   Length:90          Legacy:45   Min.   :121.0
##  1st Qu.: 4   Class :character   Novel :45   1st Qu.:180.2
##  Median : 8   Mode  :character               Median :225.5
##  Mean   : 8                                  Mean   :244.2
##  3rd Qu.:12                                  3rd Qu.:303.0
##  Max.   :15                                  Max.   :548.0
##  MentalWorkload   Session
##  Min.   : 10.00   1:30
##  1st Qu.: 55.00   2:30
##  Median : 70.00   3:30
##  Mean   : 71.74
##  3rd Qu.: 90.00
##  Max.   :130.00
```

#Mental Workload

```
Data %>%
  group_by(Tool, Session) %>%
  summarise(MentalWorkload = mean(MentalWorkload)) %>%
  spread(Session, MentalWorkload) %>%
  kable()
```

Tool

1

2

3

Legacy

87.46667

72

67.66667

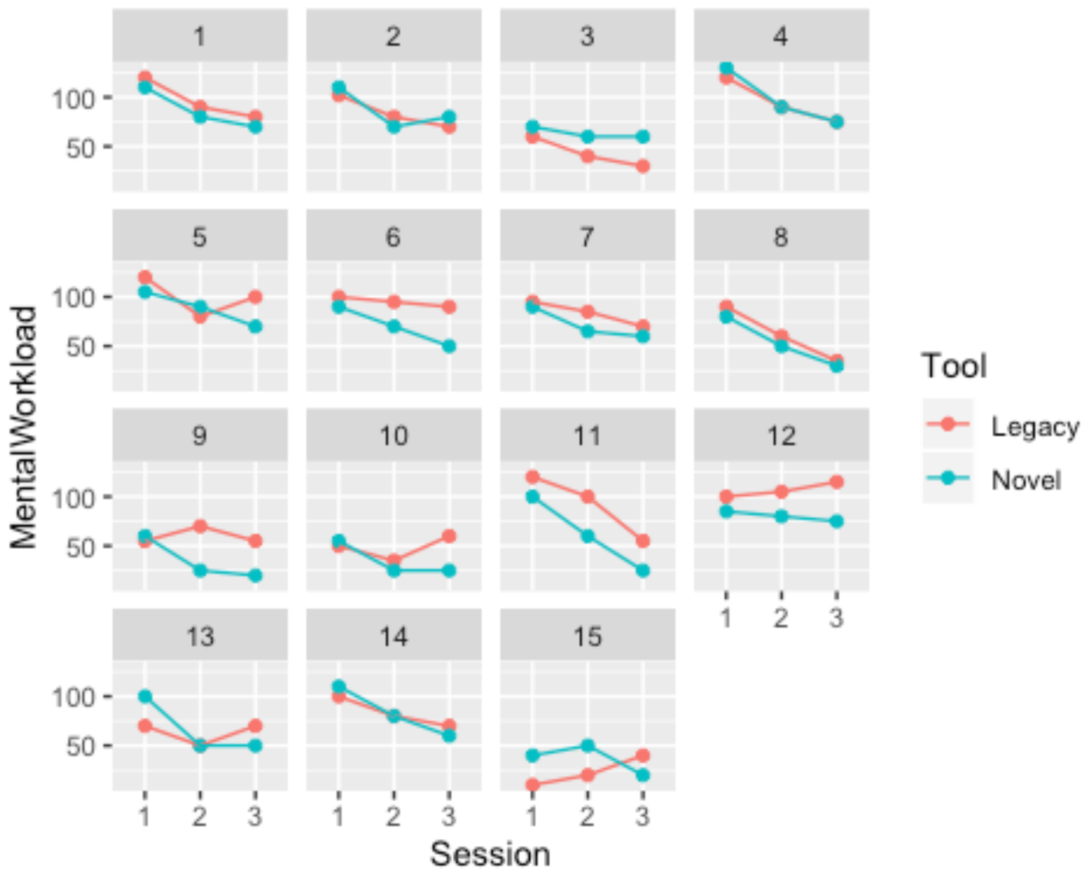
Novel

89.00000

63

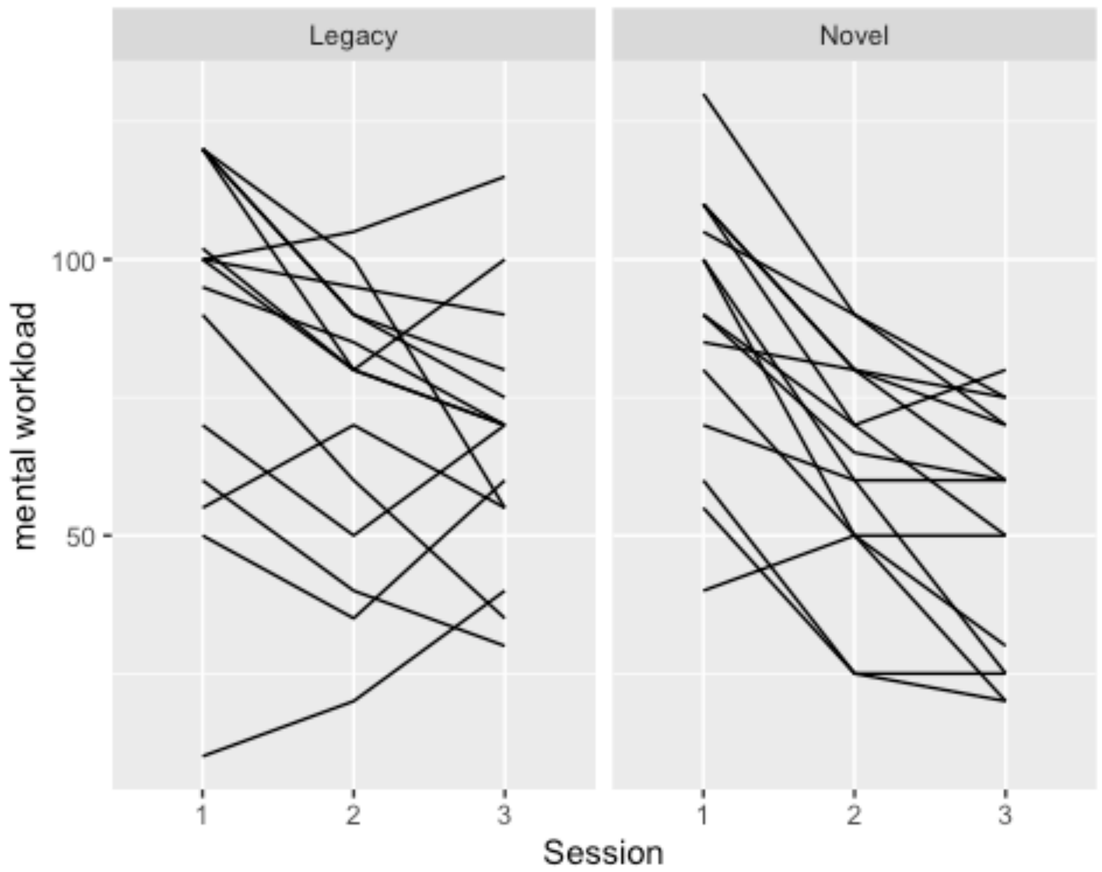
51.33333

```
Data %>%
  group_by(Participant, Tool, Session) %>%
  summarize(MentalWorkload = mean(MentalWorkload)) %>%
  ggplot(aes(x = Session, col = Tool, y = MentalWorkload)) +
  geom_point() +
  geom_line(aes(group = Tool)) +
  facet_wrap(~Participant, ncol = 4)
```




```
fig_2_mental_workload_spaghetti <-
    Data %>%
    group_by(Participant, Tool, Session) %>%
    summarize(MentalWorkload = mean(MentalWorkload)) %>%
    ggplot(aes(x = Session, y = MentalWorkload, group = Participant)) +
    geom_line() +
  ylab("mental workload") +
```

```
    facet_grid(.~Tool)

plot(fig_2_mental_workload_spaghetti)
```



##Building and running the model

```
## setting treatment contrast
contrasts(Data$Session) <- contr.treatment(3)
Data$Session <- relevel(Data$Session, 1)

glmm_MentalWorkload <- Data %>%
  mutate(MentalWorkload = as.numeric(MentalWorkload)) %>%
  MCMCglmm(fixed = MentalWorkload ~ Tool + Session +# main fixed effects
             Session:Tool,  # fixed interaction effects
           # participant level random effects
           random = ~ Participant +
             Tool:Participant + idh(Session):Participant,
           data = ., family="gaussian", # logistic family
           nitt = nitt, burnin = burnin, thin = 1) # MCMC chain length

## Warning: Unknown or uninitialised column: 'family'.

## Warning: Setting row names on a tibble is deprecated.

##
##                          MCMC iteration = 0
##
##                          MCMC iteration = 1000
##
```

```
##                          MCMC iteration = 2000
##
##                          MCMC iteration = 3000
##
##                          MCMC iteration = 4000
##
##                          MCMC iteration = 5000
##
##                          MCMC iteration = 6000
##
##                          MCMC iteration = 7000
##
##                          MCMC iteration = 8000
##
##                          MCMC iteration = 9000
##
##                          MCMC iteration = 10000
##
##                          MCMC iteration = 11000
##
##                          MCMC iteration = 12000
##
##                          MCMC iteration = 13000
##
##                          MCMC iteration = 14000
##
##                          MCMC iteration = 15000
##
##                          MCMC iteration = 16000
##
##                          MCMC iteration = 17000
##
##                          MCMC iteration = 18000
##
##                          MCMC iteration = 19000
##
##                          MCMC iteration = 20000
##
##                          MCMC iteration = 21000
##
##                          MCMC iteration = 22000
##
##                          MCMC iteration = 23000
##
##                          MCMC iteration = 24000
##
##                          MCMC iteration = 25000
##
##                          MCMC iteration = 26000
##
##                          MCMC iteration = 27000
##
##                          MCMC iteration = 28000
##
```

```
##                         MCMC iteration = 29000
##
##                         MCMC iteration = 30000
##
##                         MCMC iteration = 31000
##
##                         MCMC iteration = 32000
##
##                         MCMC iteration = 33000
##
##                         MCMC iteration = 34000
##
##                         MCMC iteration = 35000
##
##                         MCMC iteration = 36000
##
##                         MCMC iteration = 37000
##
##                         MCMC iteration = 38000
##
##                         MCMC iteration = 39000
##
##                         MCMC iteration = 40000
##
##                         MCMC iteration = 41000
##
##                         MCMC iteration = 42000
##
##                         MCMC iteration = 43000
##
##                         MCMC iteration = 44000
##
##                         MCMC iteration = 45000
##
##                         MCMC iteration = 46000
##
##                         MCMC iteration = 47000
##
##                         MCMC iteration = 48000
##
##                         MCMC iteration = 49000
##
##                         MCMC iteration = 50000
##
##                         MCMC iteration = 51000
##
##                         MCMC iteration = 52000
##
##                         MCMC iteration = 53000
##
##                         MCMC iteration = 54000
##
##                         MCMC iteration = 55000
```

```
glmm_MentalWorkload_noint <- Data %>%

  mutate(MentalWorkload = as.numeric(MentalWorkload)) %>%
  MCMCglmm(fixed = MentalWorkload ~ Tool : Session -1, # interaction effec
t only
           random = ~ Participant +
             Tool:Participant + idh(Session):Participant,
           data = ., family="gaussian", # Gaussian family
           nitt = nitt, burnin = burnin, thin = 1) # MCMC chain length

## Warning: Unknown or uninitialised column: 'family'.

## Warning: Setting row names on a tibble is deprecated.

##
##                        MCMC iteration = 0
##
##                        MCMC iteration = 1000
##
##                        MCMC iteration = 2000
##
##                        MCMC iteration = 3000
##
##                        MCMC iteration = 4000
##
##                        MCMC iteration = 5000
##
##                        MCMC iteration = 6000
##
##                        MCMC iteration = 7000
##
##                        MCMC iteration = 8000
##
##                        MCMC iteration = 9000
##
##                        MCMC iteration = 10000
##
##                        MCMC iteration = 11000
##
##                        MCMC iteration = 12000
##
##                        MCMC iteration = 13000
##
##                        MCMC iteration = 14000
##
##                        MCMC iteration = 15000
##
##                        MCMC iteration = 16000
##
##                        MCMC iteration = 17000
##
##                        MCMC iteration = 18000
##
##                        MCMC iteration = 19000
```

```
##
##                          MCMC iteration = 20000
##
##                          MCMC iteration = 21000
##
##                          MCMC iteration = 22000
##
##                          MCMC iteration = 23000
##
##                          MCMC iteration = 24000
##
##                          MCMC iteration = 25000
##
##                          MCMC iteration = 26000
##
##                          MCMC iteration = 27000
##
##                          MCMC iteration = 28000
##
##                          MCMC iteration = 29000
##
##                          MCMC iteration = 30000
##
##                          MCMC iteration = 31000
##
##                          MCMC iteration = 32000
##
##                          MCMC iteration = 33000
##
##                          MCMC iteration = 34000
##
##                          MCMC iteration = 35000
##
##                          MCMC iteration = 36000
##
##                          MCMC iteration = 37000
##
##                          MCMC iteration = 38000
##
##                          MCMC iteration = 39000
##
##                          MCMC iteration = 40000
##
##                          MCMC iteration = 41000
##
##                          MCMC iteration = 42000
##
##                          MCMC iteration = 43000
##
##                          MCMC iteration = 44000
##
##                          MCMC iteration = 45000
##
##                          MCMC iteration = 46000
```

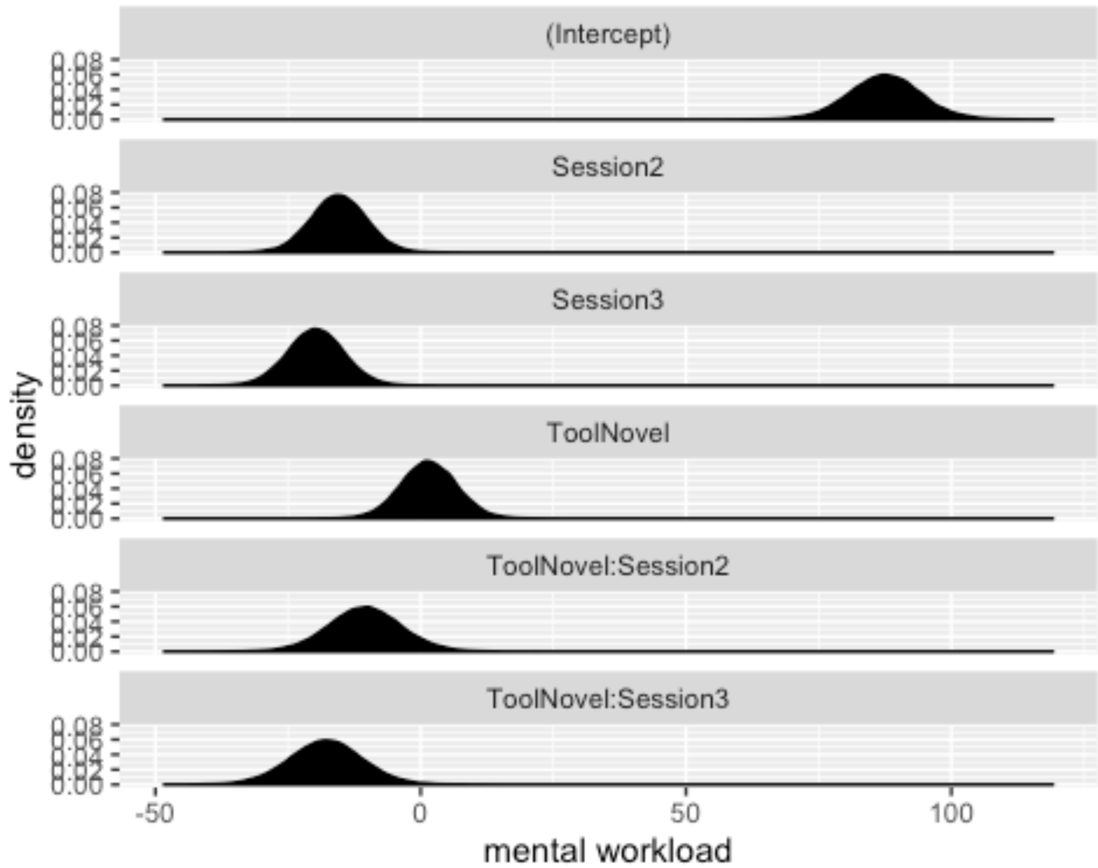
```
##
##                        MCMC iteration = 47000
##
##                        MCMC iteration = 48000
##
##                        MCMC iteration = 49000
##
##                        MCMC iteration = 50000
##
##                        MCMC iteration = 51000
##
##                        MCMC iteration = 52000
##
##                        MCMC iteration = 53000
##
##                        MCMC iteration = 54000
##
##                        MCMC iteration = 55000
```

```
summary(glmm_MentalWorkload)
```

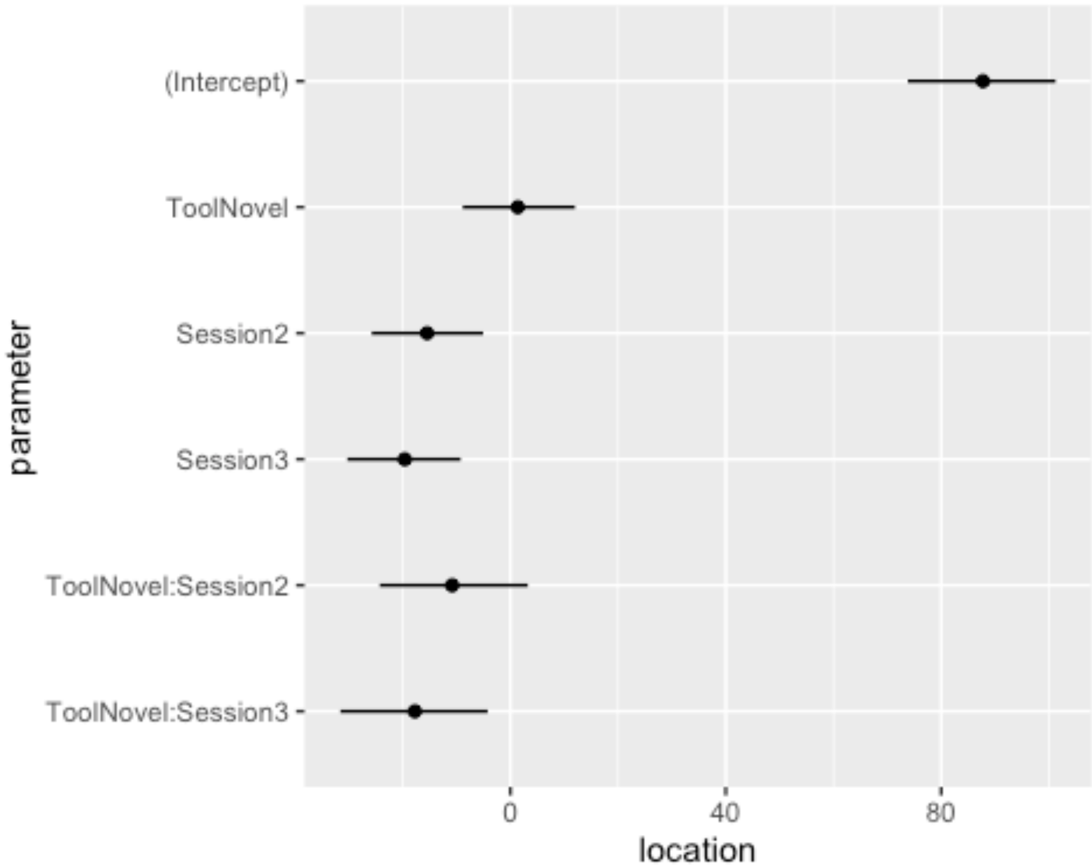
```
##
##  Iterations = 5001:55000
##  Thinning interval  = 1
##  Sample size  = 50000
##
##  DIC: 747.2678
##
##  G-structure:  ~Participant
##
##             post.mean l-95% CI u-95% CI eff.samp
## Participant     454.5    119.2    920.7     7795
##
##                ~Tool:Participant
##
##                  post.mean  l-95% CI u-95% CI eff.samp
## Tool:Participant        30 2.862e-15    135.9    298.3
##
##                ~idh(Session):Participant
##
##                      post.mean  l-95% CI u-95% CI eff.samp
## Session1.Participant   54.0120 3.870e-17  265.565    213.6
## Session2.Participant    0.4115 3.361e-17    0.283   1408.5
## Session3.Participant   15.8394 3.530e-17  114.582    453.8
##
##  R-structure:  ~units
##
##       post.mean l-95% CI u-95% CI eff.samp
## units     181.4     86.1    271.6    208.8
##
##  Location effects: MentalWorkload ~ Tool + Session + Session:Tool
##
##                    post.mean l-95% CI u-95% CI eff.samp   pMCMC
## (Intercept)           87.481   73.474  100.923    50000 < 2e-05 ***
```

```
## ToolNovel             1.531   -8.918   11.966    50000 0.77132
## Session2            -15.465  -25.791   -5.095    50000 0.00504 **
## Session3            -19.800  -30.408   -9.365    50000 0.00052 ***
## ToolNovel:Session2  -10.530  -24.514    2.978    50000 0.12804
## ToolNovel:Session3  -17.853  -31.866   -4.554    50000 0.01212 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
as.data.frame(glmm_MentalWorkload$Sol) %>%
  gather("parameter", "linear_pred") %>%
  ggplot(aes(x = linear_pred)) +
  geom_density(fill = 1) +
  facet_wrap("parameter", nrow = 6) +
  xlab("mental workload")
```




```
fig_fixef_Mental_workload <-
  fixef(glmm_MentalWorkload) %>%
  mutate(parameter = factor(parameter, levels = rev(levels(parameter)))) %
>%
  ggplot(aes(x = location, xmin = CI.025, xmax = CI.975, y = parameter)) +
  geom_point() +
  geom_errorbarh(height = 0)

## Warning: Unknown or uninitialised column: 'order'.

## Warning: Unknown or uninitialised column: 'beta'.

plot(fig_fixef_Mental_workload)
```




```
fixef(glmm_MentalWorkload, add.greek = T) %>%
  kable(digits = 3, row.names = F, align = c("l", "r", "r", "r"))

## Warning: Unknown or uninitialised column: 'order'.

## Warning: Unknown or uninitialised column: 'beta'.
```

parameter
location
CI.025
CI.975
(Intercept)
87.769
73.809
101.272
ToolNovel
1.359
-8.873
12.012
Session2
-15.475
-25.785
-5.081

Session3

-19.628

-30.304

-9.235

ToolNovel:Session2

-10.828

-24.275

3.245

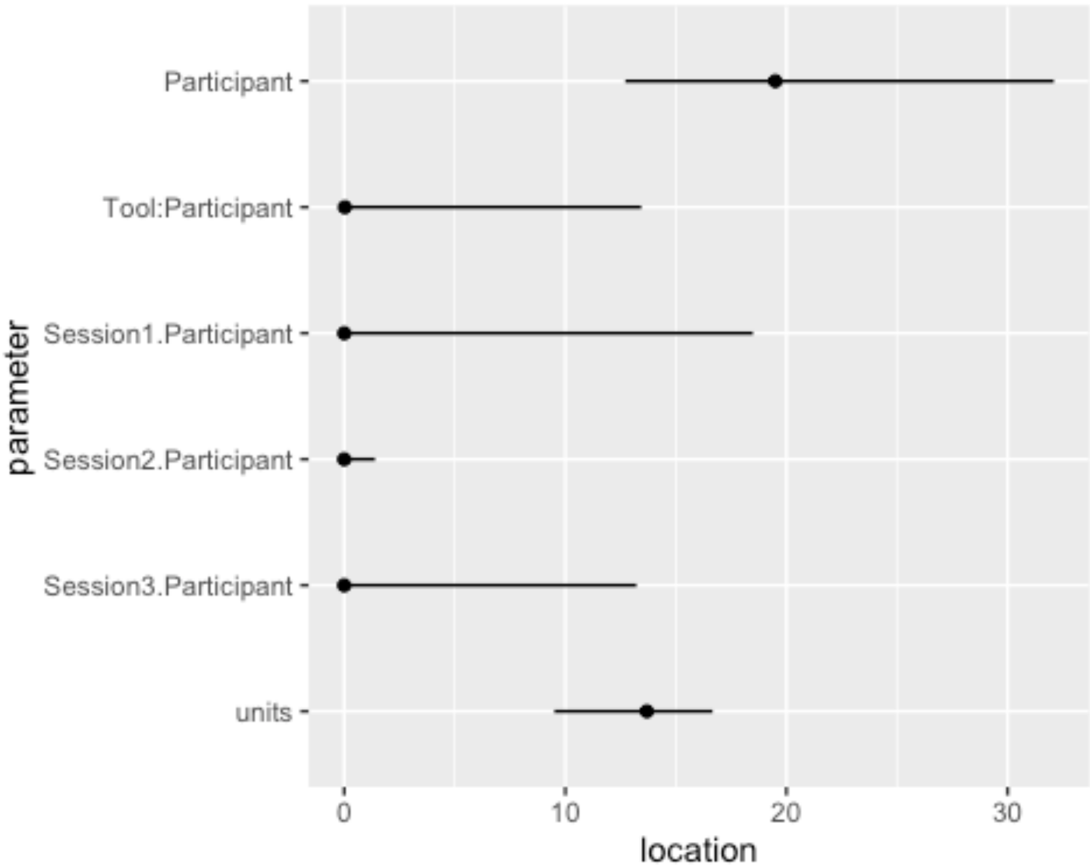
ToolNovel:Session3

-17.754

-31.589

-4.225

```
fig_ranef_mental_workload <-
ranef(glmm_MentalWorkload) %>%
  mutate(parameter = factor(parameter, levels = rev(levels(parameter)))) %
>%
  ggplot(aes(x = location, xmin = CI.025, xmax = CI.975, y = parameter)) +
  geom_point() +
  geom_errorbarh(height = 0)

## Warning: Unknown or uninitialised column: 'order'.

## Warning: `as_data_frame()` is deprecated, use `as_tibble()` (but mind t
he new semantics).
## This warning is displayed once per session.

## Warning: Unknown or uninitialised column: 'sigma'.

plot(fig_ranef_mental_workload)
```




```
ranef(glmm_MentalWorkload, add.greek = T) %>%
  kable(digits = 3, row.names = F, align = c("l", "r", "r", "r"))

## Warning: Unknown or uninitialised column: 'order'.

## Warning: Unknown or uninitialised column: 'sigma'.
```

parameter
location
CI.025
CI.975
Participant
19.496
12.711
32.095
Tool:Participant
0.028
0.000
13.440
Session1.Participant
0.003
0.000
18.488

Session2.Participant

0.000

0.000

1.401

Session3.Participant

0.000

0.000

13.223

units

13.687

9.499

16.663

#Completion Times

```
#Exploration
Data %>%

  group_by(Tool, Session) %>%
  summarise(CompletionTime = mean(CompletionTime)) %>%
  spread(Session, CompletionTime) %>%
  kable()
```

Tool

1

2

3

Legacy

349.2667

246.7333

254.3333

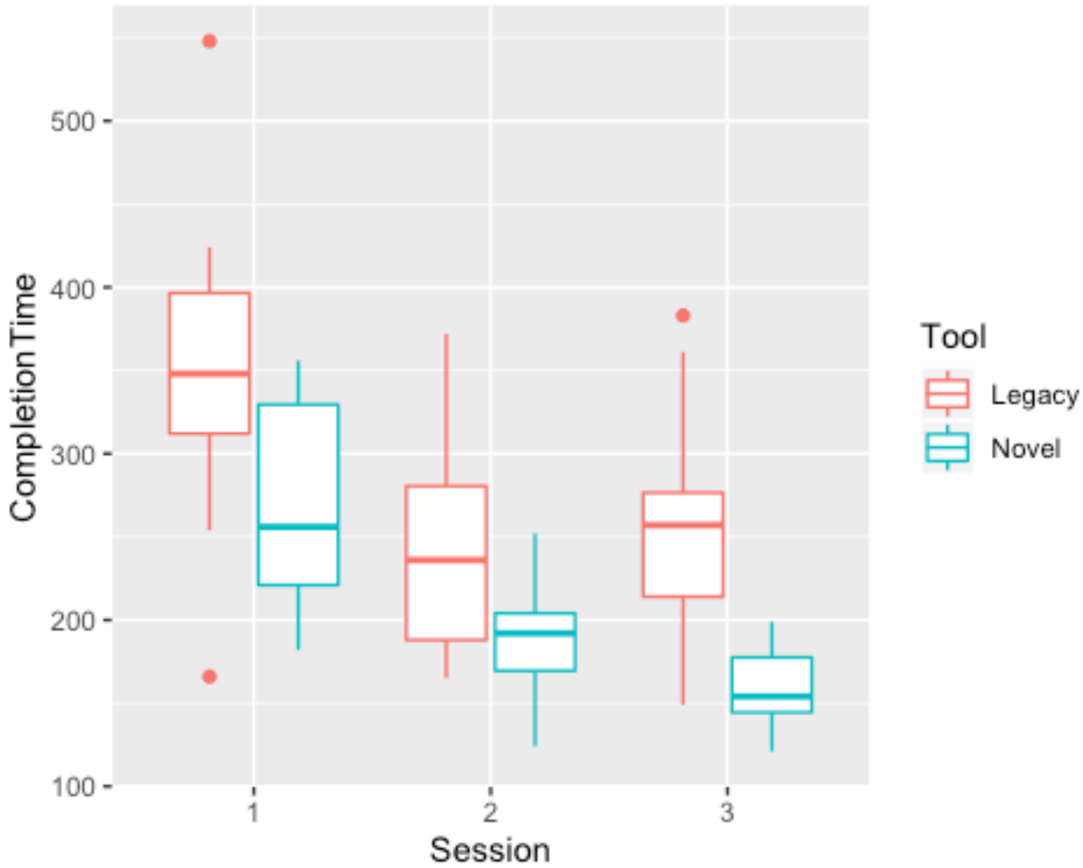
Novel

268.7333

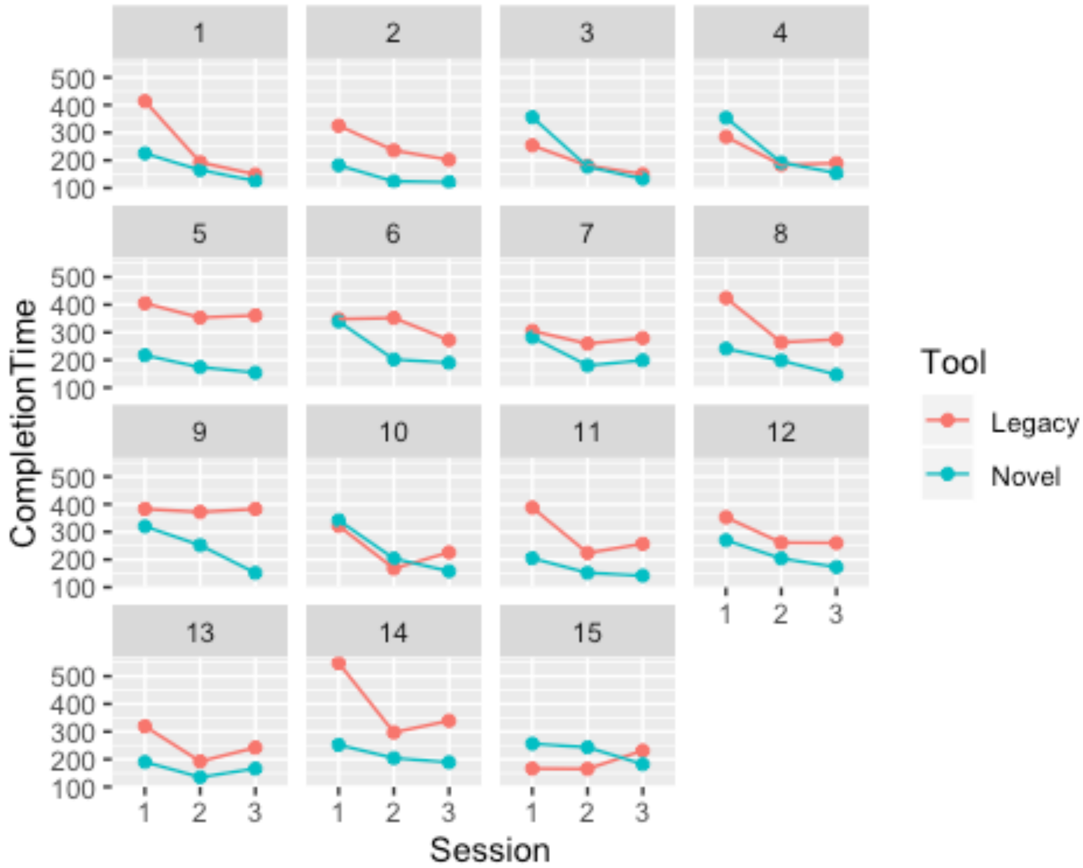
187.0667

159.0667

```
Data %>%
  ggplot(aes(y = CompletionTime,
             color = Tool,
             x = Session)) +
  geom_boxplot()
```




```
Data %>%
  group_by(Participant, Tool, Session) %>%
  summarize(CompletionTime = mean(CompletionTime)) %>%
  ggplot(aes(x = Session, col = Tool, y = CompletionTime)) +
  geom_point() +
  geom_line(aes(group = Tool)) +
  facet_wrap(~Participant, ncol = 4)
```



#Building the Model

```
contrasts(Data$Session) <- contr.treatment(3)

glmm_completionTime_noint <- Data %>%

  MCMCglmm(fixed = CompletionTime ~ Tool:Session -1,  # fixed interaction 
effects
           random =~ Participant +
             Tool:Participant + idh(Session):Participant,
           data = ., family="exponential", # exponential regression
           nitt = nitt * 2, burnin = burnin * 2, thin = 2) # MCMC chain le
ngth

## Warning: Unknown or uninitialised column: 'family'.

## Warning: Setting row names on a tibble is deprecated.

## 
##                        MCMC iteration = 0
## 
##  Acceptance ratio for liability set 1 = 0.000422
## 
##                        MCMC iteration = 1000
## 
##  Acceptance ratio for liability set 1 = 0.328244
## 
##                        MCMC iteration = 2000
## 
```

```
##  Acceptance ratio for liability set 1 = 0.131978
##
##                       MCMC iteration = 3000
##
##  Acceptance ratio for liability set 1 = 0.393878
##
##                       MCMC iteration = 4000
##
##  Acceptance ratio for liability set 1 = 0.301011
##
##                       MCMC iteration = 5000
##
##  Acceptance ratio for liability set 1 = 0.369678
##
##                       MCMC iteration = 6000
##
##  Acceptance ratio for liability set 1 = 0.332778
##
##                       MCMC iteration = 7000
##
##  Acceptance ratio for liability set 1 = 0.396322
##
##                       MCMC iteration = 8000
##
##  Acceptance ratio for liability set 1 = 0.440756
##
##                       MCMC iteration = 9000
##
##  Acceptance ratio for liability set 1 = 0.421844
##
##                       MCMC iteration = 10000
##
##  Acceptance ratio for liability set 1 = 0.410978
##
##                       MCMC iteration = 11000
##
##  Acceptance ratio for liability set 1 = 0.133156
##
##                       MCMC iteration = 12000
##
##  Acceptance ratio for liability set 1 = 0.538244
##
##                       MCMC iteration = 13000
##
##  Acceptance ratio for liability set 1 = 0.395222
##
##                       MCMC iteration = 14000
##
##  Acceptance ratio for liability set 1 = 0.233344
##
##                       MCMC iteration = 15000
##
##  Acceptance ratio for liability set 1 = 0.292156
##
```

```
##                         MCMC iteration = 16000
##
##  Acceptance ratio for liability set 1 = 0.506578
##
##                         MCMC iteration = 17000
##
##  Acceptance ratio for liability set 1 = 0.784267
##
##                         MCMC iteration = 18000
##
##  Acceptance ratio for liability set 1 = 0.775356
##
##                         MCMC iteration = 19000
##
##  Acceptance ratio for liability set 1 = 0.120978
##
##                         MCMC iteration = 20000
##
##  Acceptance ratio for liability set 1 = 0.047833
##
##                         MCMC iteration = 21000
##
##  Acceptance ratio for liability set 1 = 0.062711
##
##                         MCMC iteration = 22000
##
##  Acceptance ratio for liability set 1 = 0.307489
##
##                         MCMC iteration = 23000
##
##  Acceptance ratio for liability set 1 = 0.746911
##
##                         MCMC iteration = 24000
##
##  Acceptance ratio for liability set 1 = 0.347311
##
##                         MCMC iteration = 25000
##
##  Acceptance ratio for liability set 1 = 0.256522
##
##                         MCMC iteration = 26000
##
##  Acceptance ratio for liability set 1 = 0.758289
##
##                         MCMC iteration = 27000
##
##  Acceptance ratio for liability set 1 = 0.929456
##
##                         MCMC iteration = 28000
##
##  Acceptance ratio for liability set 1 = 0.936122
##
##                         MCMC iteration = 29000
##
```

```
##  Acceptance ratio for liability set 1 = 0.933022
##
##                        MCMC iteration = 30000
##
##  Acceptance ratio for liability set 1 = 0.800478
##
##                        MCMC iteration = 31000
##
##  Acceptance ratio for liability set 1 = 0.367778
##
##                        MCMC iteration = 32000
##
##  Acceptance ratio for liability set 1 = 0.367722
##
##                        MCMC iteration = 33000
##
##  Acceptance ratio for liability set 1 = 0.549389
##
##                        MCMC iteration = 34000
##
##  Acceptance ratio for liability set 1 = 0.458867
##
##                        MCMC iteration = 35000
##
##  Acceptance ratio for liability set 1 = 0.333244
##
##                        MCMC iteration = 36000
##
##  Acceptance ratio for liability set 1 = 0.056700
##
##                        MCMC iteration = 37000
##
##  Acceptance ratio for liability set 1 = 0.056678
##
##                        MCMC iteration = 38000
##
##  Acceptance ratio for liability set 1 = 0.128556
##
##                        MCMC iteration = 39000
##
##  Acceptance ratio for liability set 1 = 0.047456
##
##                        MCMC iteration = 40000
##
##  Acceptance ratio for liability set 1 = 0.325411
##
##                        MCMC iteration = 41000
##
##  Acceptance ratio for liability set 1 = 0.468811
##
##                        MCMC iteration = 42000
##
##  Acceptance ratio for liability set 1 = 0.170600
##
```

```
##                         MCMC iteration = 43000
##
##  Acceptance ratio for liability set 1 = 0.125356
##
##                         MCMC iteration = 44000
##
##  Acceptance ratio for liability set 1 = 0.158444
##
##                         MCMC iteration = 45000
##
##  Acceptance ratio for liability set 1 = 0.285222
##
##                         MCMC iteration = 46000
##
##  Acceptance ratio for liability set 1 = 0.441633
##
##                         MCMC iteration = 47000
##
##  Acceptance ratio for liability set 1 = 0.734622
##
##                         MCMC iteration = 48000
##
##  Acceptance ratio for liability set 1 = 0.496467
##
##                         MCMC iteration = 49000
##
##  Acceptance ratio for liability set 1 = 0.561811
##
##                         MCMC iteration = 50000
##
##  Acceptance ratio for liability set 1 = 0.464622
##
##                         MCMC iteration = 51000
##
##  Acceptance ratio for liability set 1 = 0.710778
##
##                         MCMC iteration = 52000
##
##  Acceptance ratio for liability set 1 = 0.721378
##
##                         MCMC iteration = 53000
##
##  Acceptance ratio for liability set 1 = 0.614244
##
##                         MCMC iteration = 54000
##
##  Acceptance ratio for liability set 1 = 0.554144
##
##                         MCMC iteration = 55000
##
##  Acceptance ratio for liability set 1 = 0.801600
##
##                         MCMC iteration = 56000
##
```

```
##  Acceptance ratio for liability set 1 = 0.711689
##
##                        MCMC iteration = 57000
##
##  Acceptance ratio for liability set 1 = 0.286578
##
##                        MCMC iteration = 58000
##
##  Acceptance ratio for liability set 1 = 0.251100
##
##                        MCMC iteration = 59000
##
##  Acceptance ratio for liability set 1 = 0.090689
##
##                        MCMC iteration = 60000
##
##  Acceptance ratio for liability set 1 = 0.180267
##
##                        MCMC iteration = 61000
##
##  Acceptance ratio for liability set 1 = 0.340711
##
##                        MCMC iteration = 62000
##
##  Acceptance ratio for liability set 1 = 0.020289
##
##                        MCMC iteration = 63000
##
##  Acceptance ratio for liability set 1 = 0.019400
##
##                        MCMC iteration = 64000
##
##  Acceptance ratio for liability set 1 = 0.031700
##
##                        MCMC iteration = 65000
##
##  Acceptance ratio for liability set 1 = 0.176556
##
##                        MCMC iteration = 66000
##
##  Acceptance ratio for liability set 1 = 0.207956
##
##                        MCMC iteration = 67000
##
##  Acceptance ratio for liability set 1 = 0.436389
##
##                        MCMC iteration = 68000
##
##  Acceptance ratio for liability set 1 = 0.640489
##
##                        MCMC iteration = 69000
##
##  Acceptance ratio for liability set 1 = 0.268844
##
```

```
##                          MCMC iteration = 70000
##
##  Acceptance ratio for liability set 1 = 0.074267
##
##                          MCMC iteration = 71000
##
##  Acceptance ratio for liability set 1 = 0.085756
##
##                          MCMC iteration = 72000
##
##  Acceptance ratio for liability set 1 = 0.574567
##
##                          MCMC iteration = 73000
##
##  Acceptance ratio for liability set 1 = 0.663100
##
##                          MCMC iteration = 74000
##
##  Acceptance ratio for liability set 1 = 0.654089
##
##                          MCMC iteration = 75000
##
##  Acceptance ratio for liability set 1 = 0.746278
##
##                          MCMC iteration = 76000
##
##  Acceptance ratio for liability set 1 = 0.577822
##
##                          MCMC iteration = 77000
##
##  Acceptance ratio for liability set 1 = 0.640033
##
##                          MCMC iteration = 78000
##
##  Acceptance ratio for liability set 1 = 0.426600
##
##                          MCMC iteration = 79000
##
##  Acceptance ratio for liability set 1 = 0.175567
##
##                          MCMC iteration = 80000
##
##  Acceptance ratio for liability set 1 = 0.569811
##
##                          MCMC iteration = 81000
##
##  Acceptance ratio for liability set 1 = 0.582122
##
##                          MCMC iteration = 82000
##
##  Acceptance ratio for liability set 1 = 0.739567
##
##                          MCMC iteration = 83000
##
```

```
##  Acceptance ratio for liability set 1 = 0.794889
## 
##                        MCMC iteration = 84000
## 
##  Acceptance ratio for liability set 1 = 0.747689
## 
##                        MCMC iteration = 85000
## 
##  Acceptance ratio for liability set 1 = 0.383544
## 
##                        MCMC iteration = 86000
## 
##  Acceptance ratio for liability set 1 = 0.684033
## 
##                        MCMC iteration = 87000
## 
##  Acceptance ratio for liability set 1 = 0.376222
## 
##                        MCMC iteration = 88000
## 
##  Acceptance ratio for liability set 1 = 0.450156
## 
##                        MCMC iteration = 89000
## 
##  Acceptance ratio for liability set 1 = 0.258411
## 
##                        MCMC iteration = 90000
## 
##  Acceptance ratio for liability set 1 = 0.824911
## 
##                        MCMC iteration = 91000
## 
##  Acceptance ratio for liability set 1 = 0.852622
## 
##                        MCMC iteration = 92000
## 
##  Acceptance ratio for liability set 1 = 0.765078
## 
##                        MCMC iteration = 93000
## 
##  Acceptance ratio for liability set 1 = 0.789000
## 
##                        MCMC iteration = 94000
## 
##  Acceptance ratio for liability set 1 = 0.809044
## 
##                        MCMC iteration = 95000
## 
##  Acceptance ratio for liability set 1 = 0.648589
## 
##                        MCMC iteration = 96000
## 
##  Acceptance ratio for liability set 1 = 0.608044
## 
```

```
##                         MCMC iteration = 97000
##
##  Acceptance ratio for liability set 1 = 0.250567
##
##                         MCMC iteration = 98000
##
##  Acceptance ratio for liability set 1 = 0.075833
##
##                         MCMC iteration = 99000
##
##  Acceptance ratio for liability set 1 = 0.346944
##
##                         MCMC iteration = 100000
##
##  Acceptance ratio for liability set 1 = 0.451967
##
##                         MCMC iteration = 101000
##
##  Acceptance ratio for liability set 1 = 0.058267
##
##                         MCMC iteration = 102000
##
##  Acceptance ratio for liability set 1 = 0.124422
##
##                         MCMC iteration = 103000
##
##  Acceptance ratio for liability set 1 = 0.320733
##
##                         MCMC iteration = 104000
##
##  Acceptance ratio for liability set 1 = 0.378600
##
##                         MCMC iteration = 105000
##
##  Acceptance ratio for liability set 1 = 0.736133
##
##                         MCMC iteration = 106000
##
##  Acceptance ratio for liability set 1 = 0.712700
##
##                         MCMC iteration = 107000
##
##  Acceptance ratio for liability set 1 = 0.806333
##
##                         MCMC iteration = 108000
##
##  Acceptance ratio for liability set 1 = 0.820411
##
##                         MCMC iteration = 109000
##
##  Acceptance ratio for liability set 1 = 0.554944
##
##                         MCMC iteration = 110000
```

```
##
##  Acceptance ratio for liability set 1 = 0.294967
```

```r
summary(glmm_completionTime_noint)
```

```
##
##  Iterations = 10001:109999
##  Thinning interval  = 2
##  Sample size  = 50000
##
##  DIC: 1177.836
##
##  G-structure:  ~Participant
##
##             post.mean  l-95% CI u-95% CI eff.samp
## Participant 0.0003601 4.098e-17  0.00129    350.7
##
##                ~Tool:Participant
##
##                  post.mean  l-95% CI u-95% CI eff.samp
## Tool:Participant  0.003473 8.023e-17  0.03051    12.41
##
##                ~idh(Session):Participant
##
##                      post.mean  l-95% CI  u-95% CI eff.samp
## Session1.Participant 0.0004195 4.202e-17 0.0009768   116.46
## Session2.Participant 0.0016529 3.863e-17 0.0046118    53.26
## Session3.Participant 0.0051877 4.004e-17 0.0313094    33.02
##
##  R-structure:  ~units
##
##       post.mean  l-95% CI u-95% CI eff.samp
## units  0.003396 1.007e-07  0.01684    38.91
##
##  Location effects: CompletionTime ~ Tool:Session - 1
##
##                     post.mean l-95% CI u-95% CI eff.samp  pMCMC
## ToolLegacy:Session1    -6.007   -6.357   -5.540   14.489 <2e-05 ***
## ToolNovel:Session1     -5.803   -6.179   -5.387   11.047 <2e-05 ***
## ToolLegacy:Session2    -5.576   -6.226   -5.037    5.304 <2e-05 ***
## ToolNovel:Session2     -5.309   -5.739   -4.904   12.556 <2e-05 ***
## ToolLegacy:Session3    -5.568   -5.985   -5.047   10.866 <2e-05 ***
## ToolNovel:Session3     -5.024   -5.527   -4.636   11.455 <2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

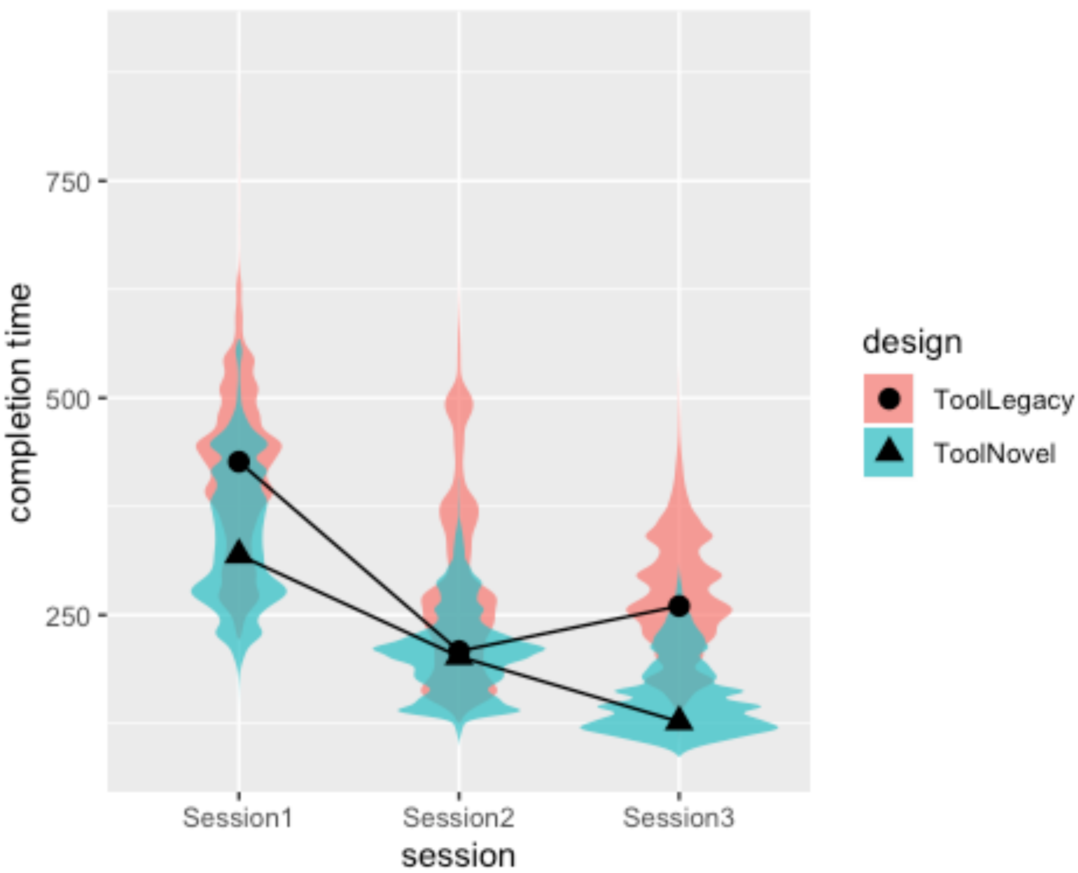
#Regression Analysis

```r
fixef(glmm_completionTime_noint) %>%
  kable(digits = 3, row.names = F, align = c("l", "r", "r", "r"))
```

```
## Warning: Unknown or uninitialised column: 'order'.

## Warning: Unknown or uninitialised column: 'beta'.
```

parameter
location
CI.025
CI.975
ToolLegacy:Session1
-6.060
-6.384
-5.553
ToolNovel:Session1
-5.764
-6.176
-5.383
ToolLegacy:Session2
-5.664
-6.222
-5.032
ToolNovel:Session2
-5.320
-5.766
-4.919
ToolLegacy:Session3
-5.572
-6.006
-5.065
ToolNovel:Session3
-4.925
-5.527
-4.636

```
#Regression Plot
fig_4_regression_plot <-
  regression_plot(glmm_completionTime_noint,
                  ylab = "completion time",
                  mean.func = function(x) exp(-x))
ylim(0,300)

## <ScaleContinuousPosition>
##  Range:
##  Limits:    0 --  300

plot(fig_4_regression_plot)
```


750
500
250
completion time
Session1
Session2
Session3
session
design
ToolLegacy
ToolNovel