Silesian University of Technology

SILESIAN UNIVERSITY OF TECHNOLOGY

FACULTY OF AUTOMATIC CONTROL, ELECTRONICS AND COMPUTER SCIENCE

PHD IN APPLIED INTEGRATIVE DATA ANALYSIS

Doctoral thesis

Research and development of a new touch-screen based inceptors design for an aircraft control

Author: Wojciech Tomasz Korek
Supervisors: Prof. Joanna Polańska, Dr Wen-Chin Li
Consultants: Dr Linghai Lu, Dr Mudassir Lone

February 2023
Gliwice, Poland

Abstract

The way of controlling things and its evolution has been with humanity from the very beginning. Over the last 100 years, the aviation, automotive, and maritime fields have seen many changes as a result of rapid technological advances. However, in terms of the user's physical input commands, only a few designs of inceptors have been applied throughout the history of those fields. This is particularly evident in the aviation industry, where the safety of the aircraft is one of the most important concerns in every design, development and utilisation stage. The research presented in this thesis explored the potential of applying alternative inceptors, such as a touchscreen or a gamepad, as controllers in aircraft flight decks. In order to test and validate these designs, a state-of-the-art engineering flight simulator was developed. In addition, the analysis aimed to determine if participants' demographic, occupational, or personal characteristics influenced their experience and performance when using these alternative controllers. The study employed an experimental design in which three different inceptors were tested in a set of simulated scenarios: a sidestick, a gamepad, and a touchscreen. Participants included both pilots and non-pilots, who were further grouped according to their characteristics (such as flight experience and video/mobile game usage). Participants' performance was measured using a variety of objective and subjective metrics in order to assess their performance, workload, situation awareness, and perceived usability of the inceptors. The results were validated using statistical analyses. It was found that the sidestick and gamepad controllers had very similar results, with an increase in performance among participants with flight experience. However, the touchscreen controller had similar, albeit lower, results among all participants, regardless of demographic, occupational, or personal characteristics. The research also highlighted the importance of considering those characteristics when evaluating the performance and experience of participants in flight simulation studies. Based on these findings, it was concluded that the touchscreen controller is not yet viable for implementation in a flight deck; however, based on a literature review and participants' opinions, there may be other potential applications where it could be used. In conclusion, this thesis provides valuable insights into the engineering flight simulator design, as well as the proposal and evaluation of alternative inceptors for aircraft control and human-computer interaction methods. The results may inform future research and development in these areas.

This page is intentionally left blank.

Acknowledgements

First of all, I would like to express my deepest gratitude to my wife, Kinga, for her unwavering support throughout this journey. Without her encouragement, love, and patience, I wouldn't have been able to complete this PhD. I am truly blessed to have her by my side.

I am extremely grateful to my family, who has been my constant source of love, support and motivation, especially my parents and Agata. If I hadn't heard "GO" from them so many times this one evening, I probably would have stayed in my hometown for the rest of my life, never realising how much I would be missing out on in this world. I would also like to thank all my friends (from all around the globe!), especially students and researchers from Dynamics, Simulation and Control Group at Cranfield University, who have been supporting and inspiring me throughout this journey.

Furthermore, I would like to extend my heartfelt thanks to Dr Wen-Chin Li, Prof. Joanna Polańska, Dr Linghai Lu, Dr Mudassir Lone, Prof. James Whidborne, Dr Saryani Asmayawati, and Dr Hafiz Ul-Asad, who were supervising me during this PhD, for their expertise, guidance and support, and especially for finding time to answer my (sometimes daft) questions, day and night.

I would also thank Peter Beecroft from Rolls-Royce for his support and approval for this research to be carried out in the Future Systems Simulator. Furthermore, I would like to express my gratitude towards Dominic Hargreaves, Dr Daniel Jenkins, and the rest of the team from DCA Design, and Dr Yuanbo Nie and Dr Andrew Mills from the University of Sheffield. Thank you for the opportunity to work with you; it was a pleasure.

Truth be told, I never planned to go abroad to study. Until the last moment, I did not intend to undertake doctoral studies. And yet, here I am. And if I could go back in time, I would do it again! And for that, I am forever grateful to everyone that made it happen. Thank you all, and remember - *be excellent to each other*!

Research was financed by the European Union through the European Social Fund (grant POWR.03.02.00-00-I029).

Contents

This page is intentionally left blank.

List of Abbreviations

ANOVA	Analysis of variance.
avg	Averaged [result].
CDG	Control display gain.
CHR	Cooper-Harper Rating [Scale].
COTS	Commercial off-the-shelf.
DOF	Degrees of freedom.
DR	Disturbance rejection.
DRH	Disturbance rejection in the horizontal [channel].
DRV	Disturbance rejection in the vertical [channel].
EASA	European Aviation Safety Agency.
EFS	Engineering flight simulator.
eVTOL	Electric vertical take-off and landing.
FAA	Federal Aviation Administration.
FBW	Fly-by-wire.
FE	Flight experience.
FG	FlightGear [Flight Simulator].
FQ	Flying qualities.
FSS	Future Systems Simulator.
GP	Gamepad.
GUI	Graphical user interface.
HF	Human factors.
HMI	Human-machine interface.
HQ	Handling qualities.
ILS	Instrument landing system.
IOS	Instructor operating station.
LD	Landing with disturbance.

LN	Landing - no disturbance.
M	Mean.
MAE	Mean absolute error.
MD	Mean difference.
MG	Mobile game(s).
NASA-TLX	NASA Task Load Index.
ns	Not significant.
PFD	Primary flight display.
PS	Performance Score.
rANOVA	Repeated-measures analysis of variance.
RMS	Root mean square [error].
SA	Situation awareness.
SART	Situational awareness rating technique.
SART-D	Situational awareness rating technique - demand.
SART-S	Situational awareness rating technique - supply.
SART-Total	Situational awareness rating technique - total [score].
SART-U	Situational awareness rating technique - understanding.
SD	Standard deviation.
SE	Standard error.
SS	Sidestick.
SUS	System usability scale.
SUS-L	System usability scale - learnability.
SUS-Total	System usability scale - total [score].
SUS-U	System usability scale - usability.
TS	Touchscreen.
UAM	Urban air mobility.
UDP	User Datagram Protocol.
UX	User experience.
VG	Video game(s).
VR	Virtual reality.

1

Introduction

1.1 Aircraft control

Inceptors are devices that allow pilots to control the aircraft's movement and functions. They consist of yokes/sidesticks, rudder pedals and engine throttle levers. Inceptors are designed to interface with the aircraft's control systems and provide the pilot with the necessary inputs to control the aircraft's speed, altitude, heading, and other parameters. These devices are a crucial aspect of the aircraft's flight control system and play a critical role in the overall safety and performance of the aircraft [122].

Handling qualities (HQ) refer to the ease and stability of controlling an aircraft, as well as the degree of confidence and predictability in the aircraft's response to pilot inputs. As an engineering field, HQ have been developing hand in hand with the discipline of flight control engineering since the beginning of aviation. This is evident from the stability and control focus of the Wright brothers [391] and the work of Norton on roll damping in the early 20th century [266]. Methods and tools for HQ assessment have had to evolve with every significant advancement in flight control design. This is exemplified by the works of McRuer & Jex [247], Hodgkinson [165], and Klyde et al. [194] that effectively chart Western evolution in the study of HQ targeting problems specific to each new flight control technology. Moreover, methods for collecting qualitative pilot feedback and commentary (such as that of Harper & Cooper [151]) have also been a fundamental aspect of HQ. Pilot workload and overall experience can also be collected using metrics such as the System usability scale (SUS) [54, 245], Situational awareness rating technique (SART) [344], or NASA-TLX (Task

Load Index) [155]. On the other hand, aircraft become more and more automatic, with pilots' role changing, as they only need to set and adjust flight parameters [152]; there is less need for a sidestick or a central yoke as they become an unnecessary weight. On top of that, inceptors have not evolved significantly throughout history [132]. It started with the Wright brothers' aeroplane, which used a stick and a cradle to control the pitch, roll and yaw of their Flyer I [391]. Shortly after that, the hip cradle evolved into the rudder pedals through the "rudder bar" (pre-1919) [205]. After several dozen years, some aircraft companies decided to replace the central stick/yoke with a sidestick [336]. All of the above were subjected to thorough research and evaluations.

So far, a traditional HQ engineer has focused on flight control architecture, stick feel systems and experimental testing (design, execution and analysis) for demonstrating performance, while, at the same time, interacting with various engineers from all disciplines to ensure the aircraft can satisfactorily perform all mission task elements. An interaction that has seen minimal change is the interface with human factors engineers in the design of inceptors. Not much has changed from the pilot's perspective regarding inceptors in the cockpit of a large commercial transport aircraft. What was used for manual control on the Airbus A300 and the Boeing 707 can be found replicated on the Airbus A350 and the Boeing 787, albeit with significant "back-end" changes in the flight control system. Today, the advancement in cockpit automation, together with the envisioned changes in the role of a pilot, has led researchers to question the suitability of existing inceptors and wonder whether better alternatives exist. The need to reduce inceptors comes from the route for autonomous control in the automotive industry [207]. Inceptors do not need to take up as much space, especially with all the machinery under the deck in contemporary aircraft. It is possible that the change of inceptors will reduce the aircraft's total weight. This, in turn, might reduce the production costs and fuel burn, and increase the aircraft's potential lifespan [163]. Other aspects include the inceptor's usability and the pilot's performance. Recent work in the field has explored designing a control system that helps pilots with limited experience to fly safely [355], with emphasis on aspects of learnability and safety. Other studies focused solely on comparing existing inceptors (sidestick and central yoke). A series of experimental research has shown that type of the inceptor is one of the numerous variables in the pilot-vehicle system, and its parameters like position (central/side), displacement/force, stiffness and damping can influence pilot performance [113, 114, 395]. To date, much of this exploration has been limited to traditional inceptors such as sidesticks and control columns. Therefore, it is worth considering if **an introduc-**

tion of alternative inceptors as flight controllers, including a touchscreen, can potentially improve the pilot's performance. Furthermore, there is a growing trend to seek alternative control methods for urban air mobility (UAM) vehicles, with researchers investigating sidesticks, mouse-like devices, and gamepads [129, 231, 306]. With such directions in aviation, a state-of-the-art flight simulator environment is needed to accommodate research into alternative inceptors.

1.2 Flight simulation

The safety-critical nature of aviation has been acknowledged as one of the major reasons for the limited changes in aircraft inceptors [114]. Due to the high-stakes environment of aviation, any new concepts must undergo a series of rigorous tests and adhere to strict standards before being deemed airworthy. The consequences of an airborne failure in an aircraft are severe, making it crucial to ensure the safety of the passengers and crew. However, modern technology and research facilities, such as engineering flight simulators, allow for safe testing of new ideas in controlled environments [11].

Engineering flight simulators (EFSs) can be a powerful tool for novel research, allowing for the safe and controlled evaluation of alternative inceptor designs. However, the traditional approach to flight simulation research can be time-consuming and resource-intensive, making it difficult to explore a wide range of control methods and designs, such as the introduction of touchscreen technologies in a flight deck. Therefore, a second goal of this thesis was the development of the Future Systems Simulator (FSS) - a simulator which brings an invaluable contribution to the science of investigating human factors, human-machine interface, and novel technologies in aircraft, among many other fields. Moreover, a unique, user-centred technique was utilised when designing the cockpit. This method is proposed as a new principle for researchers and companies designing an EFS. It is believed that **this novel flight simulator can be utilised to streamline the research and validate the results of radically different control methods in an aircraft**. This will provide valuable insights into the potential benefits and drawbacks of alternative inceptors, helping to guide the development of next-generation flight control systems.

1.3 Human factors

Human factors (HF), also known as ergonomics, is a multidisciplinary field of study that focuses on understanding how people interact with the systems and technologies they use and how to design and optimise these systems and technologies to best support human performance and well-being. HF research takes into account a wide range of aspects that can affect human performance, including physical, cognitive, social, and organisational characteristics. The goal of HF research is to enhance the safety and efficiency of human-system interactions by identifying and addressing potential sources of human error and designing systems and technologies that are easy to use, understand, and control [309].

Human-centred design is an HF technique that focuses on understanding the needs and limitations of the end user. It takes into account the user's perspective and incorporates their feedback throughout the design process. The goal is to create products, systems, or services tailored to the user's needs, and are easy to use, understand and interact with [84].

Demographic, occupational and personal characteristics are considered among the many factors affecting the user's behaviour. For example, demographic and personal characteristics such as cultural background, interests or attitude towards technology may affect how a person perceives and interacts with a product [64–66, 200, 375], while occupational characteristics such as experience, training, and role may affect how a person uses a product in their work [9, 198]. Therefore, these characteristics should be considered as an important part of the user-centred design process.

The third aim of this study was to investigate the potential impact of demographic, occupational, and personal characteristics on pilots' subjective experience and objective performance in a flight simulator. The hypothesis was that **those characteristics would significantly affect the subjective experience and objective performance in the flight simulator.** This hypothesis was based on the idea that factors such as age, attitude towards touchscreens in a flight deck, video game usage, and flight experience may play a role in determining a pilot's ability to adapt to and operate new inceptor technologies. The study used a combination of subjective and objective measures to evaluate this effect, with the goal of providing insights that can inform the design and development of new inceptor technologies.

1.4 Motivation and hypotheses

The evolution of aircraft technology has resulted in the ongoing requirement for modifications and upgrades to be made in order to comply with new standards, regulations, or even participation in an "arms race". This has been demonstrated through the history of the Boeing B-52, which remains in service to this day despite being produced in the early 1950s. However, significant changes such as the replacement of the entire cockpit have been necessary over the course of its service [195]. Historically, the process of making such changes was faster from a certification perspective. For example, during the Cold War, changes to aircraft were able to be made more rapidly due to the need to keep pace with the latest military technology [117]. This is in contrast to more recent times, where the process of updating and modifying aircraft has been subject to significant regulatory constraints. However, the advent of "glass cockpits" and advancements in touchscreen technology have the potential to provide a solution to this problem, offering a high degree of modularity and flexibility in aircraft cockpit design [153, 367]. The use of modular and flexible touchscreen monitors and panels presents a range of possibilities for the integration of such technology not only in the production of new aircraft but also in the upgrading of existing ones, thus reducing costs and minimising the risk of non-airworthiness.

Inceptors design for aircraft control is a topic that is often omitted by researchers, despite some potential advantages. With the rise of touchscreen technology being introduced in the aircraft's cockpits and growing interest in urban air mobility, the author observed the possibility of exploring and introducing new means of controlling the aircraft. This thesis introduced two alternative inceptors: a gamepad, usually found as a controller for video game consoles, and a touchscreen, adapted from a thumbstick commonly used in mobile games. Both were compared to a conventional sidestick. Gamepad has a remarkably ergonomic design, which has been perfected for years. Researchers have already used it in other areas than video gaming [49, 67, 166, 237, 374]; none of them, however, attempted to integrate it with a future flight deck. Touchscreen technology, on the other hand, was investigated widely in terms of interfacing with various aircraft systems, such as navigation and flight management, but never to replace the physical inceptor [46, 153, 212]. Based on those findings, the first hypothesis can be defined as:

Hypothesis 1 ($\mathbf{H_1}$): *The introduction of alternative inceptors as flight controllers, including a touchscreen, can potentially improve the pilot's performance.*

Engineering flight simulation provides a critical environment for such investigations [11]. Therefore, it is important to have a flexible, reconfigurable flight simulator that allows research not restricted by the physical characteristics of existing cockpits. For that purpose, a state-of-the-art EFS was built to investigate the following hypothesis:

Hypothesis 2 ($\mathbf{H_2}$): *A novel engineering flight simulator helps to streamline the research and validate the results of radically different control methods in an aircraft.*

Human factors play a major role in the usability analyses of any system. It is important to recognise the impact of demographic, occupational, and personal factors in the research and development of new technologies in aviation. The adoption of fully customisable touchscreen technology in flight decks (and other aviation departments such as air traffic control) could help towards safer environments for people with different personalities and cultural backgrounds [64–66, 375], as it would allow them to adjust systems according to their needs and habits. From that, a third hypothesis was defined:

Hypothesis 3 ($\mathbf{H_3}$): *Demographic, occupational, and personal characteristics have a significant effect on the subjective experience and objective performance in the flight simulator.*

1.5 Thesis outline

This thesis is structured as follows: Chapter 2 expands the aircraft control, flight simulation, and human factors topics and links to the hypotheses introduced in Chapter 1 by presenting a literature review on the following:

- simulation origins, and a history of flight simulation (links to hypothesis (H_2)),
- human-system interaction - a review of interfaces used in other areas than fixed-wing aviation (links to hypothesis (H_1)),
- evolution and contemporary research of the flight deck inceptor (links to hypothesis (H_1)),
- touchscreen technology - a history and modern applications in the flight deck (links to hypotheses (H_1) and (H_3)), and
- urban air mobility – electric vertical take-off and landing (eVTOL) vehicles (links to hypothesis (H_1)).

Chapter 3 presents the background, motivation, design philosophy, architecture and the author's contribution to the development of a state-of-the-art engineering flight simulator at Cranfield University, the Future Systems Simulator. The following Chapter 4 presents a methodology used in the data collection process. It consists of the experimental setup, participants' information, inceptors and tasks presentation, trial procedure and collected data details, along with a definition of the subjective rating scales used in this research. Beginning of Chapter 5 summarises findings from the pilot study published by the author [197]. Next, it presents how the collected data was organised, processed, and filtered. After that, it presents the analyses of significant factors affecting the results from each inceptor, presenting the findings at the end. Finally, Chapter 6 summarises the trial outcomes and work done during this research. Furthermore, it links the outcomes to the hypotheses, highlights the contributions to knowledge and proposes further paths for similar research, while reminding of existing limitations.

2

Literature review

2.1 Simulation origins

It is hard to specify when the concept of "simulation" began. In the medical field, it started in Middle Ages, when doctors used dead animals to train their surgical skills; then, the first medical mannequin, "phantom", was made in the eighteenth Century [185]; in medical education, simulation appeared in the 1960s as part of resuscitation, anaesthetic and clinical skills training [296]. One of the first "simulators" as such is considered to be a game of chess, defined around the sixth century, which was supposed to simulate battlefield tactics [262]. Another approach in the search for simulation origins is mathematical modelling. Mathematicians have been trying to describe things and phenomena that surrounded them using equations for years, which can be traced back to the seventeenth century and the invention of "modern" calculus by Isaac Newton and Gottfried Wilhelm Leibniz [37]; or even to the Renaissance, and *De revolutionibus orbium coelestium* by Nicolaus Copernicus [287]. Back then, it was just calculations and observations, resulting in a set of numbers or hand-drawn figures that only scholars could understand. Today, theoretical numerical analyses still exist and play a significant role in all areas of life; however, in terms of skills training (such as for a pilot, driver, heavy equipment operator or surgeon), no theory will give as much experience as physically being in a simulated cockpit, car, or surgery room. Simulated environments can introduce the user to various dangerous situations, at the same time being entirely safe for them.

2.2 A century of flight simulation

2.2.1 Past

Nowadays, it is normal for a pilot to have thousands of hours spent in large, commercial training flight simulators before getting aboard a real plane. Almost one hundred years ago, such training was not possible, so the pilots had to earn their piloting skills in real flight, which often caused them to fly in dangerous conditions without proper training. This has led to many fatal accidents [86] and was acknowledged by many engineers and aviation experts [293]. The first attempts to build something that one could call a "flight simulator" were done just seven years after the first Wright brothers plane. In 1910, Haward came up with possibly the first definition of a flight simulator: "a device which will enable the novice to obtain a clear conception of the workings of the control of an aeroplane, and of the conditions existent in the air, without any risk personally or otherwise" [158] (as reported by Rolfe & Staples [294]). The same year has brought several approaches to this problem: Wright brothers' *Kiwi Bird*, *The Sanders Teacher*, Billing's *Oscillator*, and *Antoinette Trainer*, among others. Kiwi Bird was an unflyable *Wright Type B Flyer*, which the Wright brothers put on a special pedestal with an electric motor-driven base [149]. The Sanders Teacher and Billing's Oscillator were built similarly, but they used the power of the wind to simulate flight conditions and turbulence; the aim of those was to maintain a steady "flight" - equilibrium. The difference between The Sanders Teacher and Oscillator was that the first one used real aircraft parts (it was based on Sander's *Type 2* aeroplane, while the Oscillator was built purely for balance-maintaining purposes, without resemblance to any actual aircraft [48, 157, 158]. Antoinette Trainer, also known as *Tonneau Antoinette* ("Antoinette barrel", shown in Fig. 2.1), was just a device consisting of two halves of a barrel with ropes which people pulled in different directions to

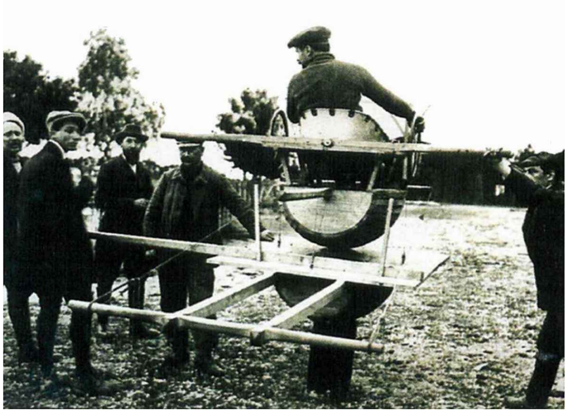


Figure 2.1: *Tonneau Antoinette.* Reproduced from Antoinette aircraft company under Free Art License [19].

simulate the movement. Those simulators were helpful in pilot training; however, they were not going to teach the pilot how to fly; it was more to accustom them to various situations they might experience in flight [275]. Other examples of "pre-simulators" are *Bleriot Simulator* (1909-1915), where a pilot was blindfolded; this device was used for sensory studies and flight personnel selection [157]; Sparmann's *Aiglon* (1911), which worked in a wind-based way, similar to The Sanders Teacher [330] (as reported by Bolton [48]); Breese *Penguin* (1917-1918), which was an actual aircraft, albeit with wings too small and engine too weak for actual flight. In this simulator, students were able to experience near-flight situations [88]; *Ruggles Orientator* (1917), which was a 3-axes motion device capable of moving in 360 degrees along each axis. It additionally was placed on a platform with wheels, adding additional vertical movement [18]; Beech's *Terra Tutor* (1917) was USA's answer to Billing's Oscillator, but, instead of wind, it used air pipes to move the device [14] (as reported by Bolton [48]). Another device, which is considered one of the first flight simulators, was invented during World War I. It was a fixed-base training rig developed to train aeroplane machine gunners aiming down moving targets, like enemy aircraft. The trainees needed to learn to aim ahead (which was called "deflection shooting") and to use special ring sight [91].

The breakthrough in flight simulation technology is attributed to Edward A. Link. The 1929 *Link Trainer*, or the "Blue Box", as it was affectionately called by the student pilots, was the first flight simulator to accommodate a 3-axis motion moving platform with flight cockpit instruments, along with an instructor station which plotted the aircraft's simulated dynamics. The platform, shown in Fig. 2.2, enabled pitch, roll, and yaw movement and was controlled by pressurised air [226]. At first, it did not gain much popularity - until 1934, when, after the "Air Mail Fiasco" [86], The Army Air Force ordered six Link Trainers, which gave Link wider recognition and is considered the start of the flight simulation industry [18, 149, 382]. It had most of the modern flight simulator components - moving platform, radio communication between pilot and instructor, instructor's station, data logging and, as one of the first flight simulators, working cockpit instruments [293]. All those elements attributed to success - Link's trainer was in use for a long time: until the late 1950s, over 10,000 units were produced.

Other notable simulators that emerged in the 1930s and 1940s were Dehmel's *Trainer* and the Travis *Aerostructor* - with fixed base, they are considered the first electrical flight simulators [275]. In 1948, Curtiss-Wright provided Boeing 377 *Stratocruiser* simulator for Pan American Airways, which became the first airline to own a flight simulator. The technological advances in digital computers in the 1950s and

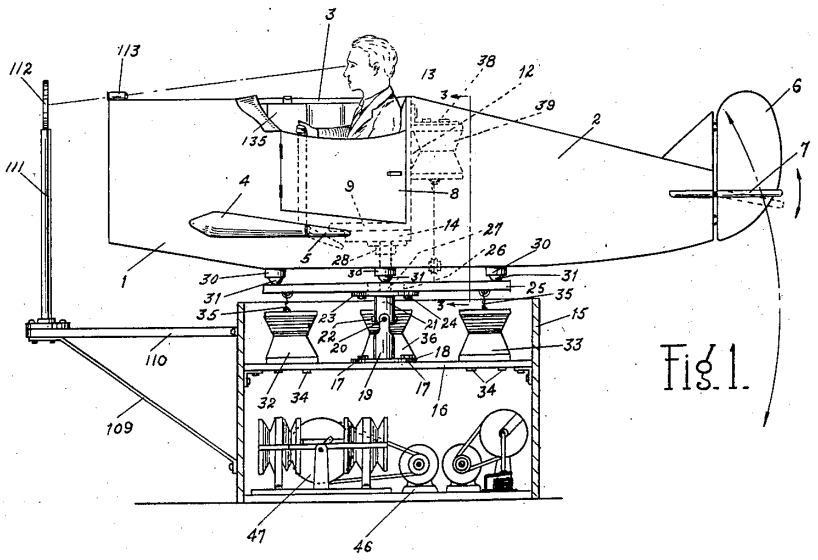


Figure 2.2: A drawing of *Link Trainer*. Reproduced from Link's patent under Public Domain license [226].

1960s brought new capabilities to flight simulators - real-time digital simulation became possible [26, 275].

However, it was still missing one important aspect of flight simulation - a visual, "out-of-the-window" system. First attempts in adding the visual cues consisted of utilising the point-light source projection method [162] and then Close-Circuit Television (CCTV) technology, which projected the rolling tape with a picture of terrain onto the screen; the speed of the terrain and position of the camera changed based on pilot's inputs. Monochromatic at first, then in colour in 1962 [161]. In 1971, the first Computer Generated Image (CGI) visual system for flight simulation was developed by McDonnell-Douglas Corporation in their *Vital II* simulator [26]. The visual system based on projectors, nowadays often used with curved or spherical screens, is used up to date [275]. In the 1960s, active inceptors began to arise. The flight instrument simulator (*FISIM 1A*) at the Royal Aircraft Establishment (RAE) at Farnborough featured pneumatic feedback in the "force-and-aft direction"; however, the movement was limited by the motor servos up to only 1.5 cycles per second [282].

2.2.2 Present

Nowadays, flight simulators can be categorised based on their purpose: research (engineering), training (commercial) or entertainment. Commercial flight simulators, which emerged from the trainers described above, can be further categorised. Those categories depend on organisations or countries' regulations; most notable are those from US Federal Aviation Administration (FAA) and European Aviation Safety Agency (EASA). Both slightly differ from each other, although for full flight simulators, the scale is similar, with level "A" being the most basic with 3 degrees of freedom (DOF) motion system, up to level "D" being the most realistic with 6-DOF motion and vibration systems [3, 118]. Up to the mid-1950s, all simulators focused strictly on pilot training. Those were considered "commercial" or "training" flight simulators. The rise in computing power and mathematical modelling allowed for more realistic simulation, even of aircraft that were not built yet. One of the first "acceptable" digital reconstructions of actual events was done by North American Aviation after the crash of the F-100 *Super Sabre* in 1954 [130]. This allowed for a more detailed analysis of the incident in order to prevent it from happening in the future. Since then, flight simulation has become present in aircraft design and testing process - for example, it was a major factor in designing the North American X-15 rocket-powered jet [327] (as reported by Baarspul [26]) and Concorde [214]. Some of the first notable engineering flight simulators (EFS) include RAE Farnborough *FISIM 2* [282], RAE Bedford *Advanced Flight Simulator* [357], National Aerospace Laboratory moving base research flight simulator [179] and NASA-Ames *Flight Simulator for Advanced Aircraft* (FSAA) [402]. In the 1980s, NASA-Ames and Lockheed-Georgia Company developed the *Advanced Concepts Flight Simulator* (ACFS). It was built using state-of-the-art technology advances at the time, including the cathode-ray tube (CRT) cockpit displays controlled by touch-screen and voice commands [68]. With many upgrades throughout the years, it is probably one of the most recognisable EFS facilities, and it has facilitated many various research studies [44, 232, 244]. It is worth noting that there is also another type of flight simulator, In-flight simulators. However, they are out of the scope of this review; more information can be found in the article by Baarspul in Progress in Aerospace Sciences [26].

2.2.3 Future

Nowadays, commercial flight simulation advances can be seen in two ways: the first one is more conventional and includes further improvements to the fixed- and moving-base simulators so they are more reliable and cost-effective. The second approach is more extreme and involves getting rid of the cockpit and projection system, two components that played a major role in flight simulators for many decades. As a result of the increasing access and decreasing costs of virtual reality (VR), there is a rise of interest in exploring VR in flight simulation. This will reduce the need for big, heavy, and, most importantly, costly cockpits; the downside of that is there will be no tactile feedback on the actions performed by the pilot [270], and it can lead to so-called "cybersickness" [60, 108, 181]. It can be used, however, as an addition to normal flight simulator training, as it can effectively decrease the flight simulator time without compromising the learnability of skills and cockpit procedures. Moreover, recently there have been studies on using mixed reality (MR) or augmented reality (AR) [39, 83, 220, 350, 362] with the existing flight simulators, which were extensively reviewed by Cross et al. [90]. Some aircraft manufacturers like Airbus [5] and Boeing [47] have already introduced VR in their pilot training. External companies, for example, Visionary Training Resources [372], that offer such training capabilities are also emerging. And yet, touchscreen controls can give more haptic feedback than virtual reality flight simulators (VRFS). It is worth noting that mixed-VRFS also exist; however, such simulators require physical cues to be mounted and calibrated with the virtual environment, and the more "fixed" cues they have, the less flexible they are [270].

2.2.4 Summary

The first physical flight simulators emerged over 100 years ago, not long after the Wright brothers' plane. The "real" simulation, however, dates back to Link's Trainer, developed in the 1920s and 1930s. The "Air Mail Fiasco" has proven that it is necessary to implement simulation into education and training processes [86]. Fig. 2.3 summarises the first 100 years of flight simulation with some of the most important events in its history. Throughout the years, safety aspects in the aviation industry have switched from "post-incident/accident" investigations and avoidance of certain situations in the future (reactive approach) to reacting to potential risks before they happen (proactive approach). Flight simulators can play a major role in such methods as they offer a safe environment to test dangerous or risky situations and scenarios [99]. Moreover, this event has given a significant rise in simulation interest in other fields as well, such as medical [185] or automotive.

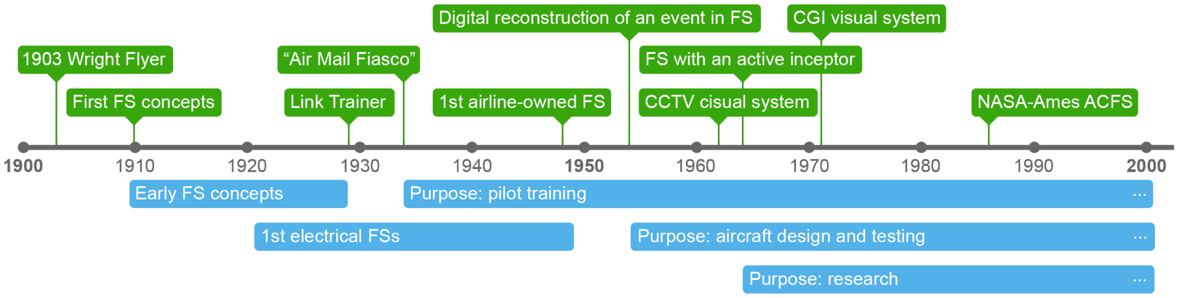


Figure 2.3: Timeline of the most important events in the first 100 years of flight simulation. FS - flight simulator.

Over the last hundred years, along with aviation technology, flight simulation has been subjected to great advancements. The graphic quality of the visual system got more realistic, the interaction between the pilot and the cockpit became more important, and the aircraft models, simulated in real-time, got much more credible. All of the existing flight simulators are based on the already existing aircraft and flight decks. To this date, EFSs, even the NASA-Ames ACFS, have been representing existing aircraft environments and systems. They had physical levers, switches and knobs, as can be found in real cockpits. Any major changes to the layout were either impossible or very costly. The *Future Systems Simulator* (FSS), described in Chapter 3, aims to provide a platform for rapid (and long-term as well) prototyping, giving the opportunity to show current and future pilots radically different concepts of flight deck design, information provision and flight experience while at the same time focusing on safety and reliability of all the aircraft systems, addressing the hypothesis (H_2).

2.3 Human-system interaction

How do humans control things around them? Who was the first person to think that a boat or a car could be steered by moving a big wheel in front of us, or come up with the idea that, when a specific point is touched on a mobile phone screen (or later a tablet), something expected will happen without the need to use a directional pad or mouse? The way of controlling things and their evolution has been with humanity from the very beginning. Some inventions were so well-thought that they stayed for hundreds of years and did not change much. Some have been in development through many years and still are not perfect - there is still much to discover. It is difficult to find specific information on when the first ”controller” or ”inceptor” was invented and how it was used, as it depends on the scientific area and the term’s definition. One can look for the first usage of the word, but that is out of the focus of this study. This section covers various control methods in different areas of life, and based on that literature review, the decision on which inceptors can be introduced in the flight simulator is made, addressing the hypothesis (H_1). Aspects of inceptor evolution in aviation were discussed in Section 2.4. Here, the author presents other areas where this topic is also applicable.

2.3.1 Video games and simulation

Controllers in video games play a major role. The gaming industry is one of the biggest entertainment industries in the world, recently even surpassing the movie and sports industry in annual revenue [361, 388] – and it is not only because of the Covid-19 pandemic. Coronavirus just sped this process up [256]. Contrary to watching movies or sports events and reading books, video games need constant interaction from the user. From the very beginning, video game controllers evolved in parallel with the software and hardware: starting with just two knobs (called ”paddles”) in one of the first commercial gaming systems, Magnavox *Odyssey* and Atari’s *Pong* (1972); through ”one-button joysticks” (with the idea of ”joystick” adapted from aviation technology) in Atari *2600* (1977); vast selection of gamepads (first appearing in Nintendo’s *Family Computer*, 1983); and then gamepads known from Sega, Xbox and PlayStation); up to motion tracking systems (Nintendo Wii, Xbox Kinect, PlayStation Move) and touchscreen controllers as found on smartphones and Nintendo DS/3DS.

The first commercial gaming system was Magnavox Odyssey. Although released in 1972, the first prototype was built by Baer in 1968. It was something resembling an old table tennis game, where the player (visualised by a white dot on the screen) could move in 4 directions using two potentiometers. 4 years later, it was released as the first commercial video gaming system. In the same year, Atari, inspired by Baer’s invention, released ”Pong”, which was also a table tennis game, albeit with players moving only vertically using just one knob (the device had two knobs – one for each player) [193]. In 1977, Atari introduced *Video Computer System* (later called Atari 2600), which had a joystick and paddle controller, although it was not the first time the joystick was used. Joystick as a device has been present since the beginning of the aviation industry, with the first usage of the term dating back to 1909 [384]. In the gaming industry, they were first used in the 1973 *Astro Race* arcade video game, manufactured by Taito company [173]. The joystick was a popular controller choice in a number of consoles released up to the mid-80s. Even though they were surpassed by gamepads and later keyboards as the main controller in video game systems, they are still popular today, mainly in desktop flight simulators. The first gamepad was proposed by Nintendo's *Family Computer* (Famicom) in 1983. It had a cross-shaped directional pad ("d-pad") and four extra buttons. North American version of that console, Nintendo *Entertainment System* (NES) (1985), retained the same idea for the controller. Gamepad design was perfected throughout the years, with notable approaches being attributed to Sega Mega Drive (1988-1999), Sony PlayStation (1994-2020), Nintendo 64 (1996), Nintendo GameCube (2001-2002) and Microsoft Xbox (2001-2020)[1] [370]. With numerous studies in areas like ergonomics, human factors or system usability, and years of improving the design, Microsoft's Xbox and Sony's PlayStation are current pioneers in terms of state-of-the-art gamepad design.

The turn of the 20th and 21st centuries has brought a number of innovations in video games control: in 1994, Thrustmaster released the first steering wheel controller for PC racing video games, *Formula T1* [319], although the steering wheel as a controller in arcade video system was used as early as 1981 [168]. In 2006, Nintendo introduced Wii *Remote*, revolutionising user input by tracking the motion of their arms. In this case, the user had to hold the device; in 2010, Microsoft presented *Kinect* – a device for Xbox consoles with a camera that could track the user's entire body movements and gestures, as well as voice commands. In the same year, PlayStation, being Xbox's main competitor, released a motion-sensing controller as well. However, unlike Kinect, PlayStation's *Move* required a handheld device, similar to Wii's Remote. Kinect has gained a notable interest in the research community. There are several studies utilising this controller [399]. Some researchers have also proposed touchscreen-based controllers by utilising mobile phones as inceptors [238, 358].

Parallel to stationary gaming systems with separate controllers, a field of handheld game consoles emerged in the 1970s. The concept was to have a device that could be played anywhere instead of only in front of the TV or other screen. With

---

[1] Years in brackets indicate first and latest models produced.

origins in electro-mechanical devices, offering just a single game at first (Cragstan's *Periscope-Firing Range*, produced in the 1960s [257]; Waco's *Electronic Tic-Tac-Toe*, 1972 [100]; Mattel's *Auto Race*, 1976 [230]), and through interchangeable cartridges (Milton Bradley's *Microvision*, 1979 [323]), handhelds gained widespread popularity after the release of first *Game Boy* by Nintendo in 1989 [334]. Since then, Game Boy's successors are dominating the handheld market. The first Game Boy had a similar set of buttons as Famicom's or NES's gamepad: a cross-shaped d-pad and a number of buttons. Nintendo had not changed this setup until 2004 when they included a touchscreen controller (although the first handheld console to feature touchscreen control was the Tiger Electronics *Game.com* console from 1997 [96]). Nintendo 3DS from 2011 featured a trackpad, and Nintendo Switch from 2017 included analogue sticks. Some of the notable competitors include Sony *PlayStation Portable* (PSP) (2004), which included flat, analogue thumbstick controller years before Nintendo; the Second handheld from Sony, PlayStation Vita (2011), which had a touchscreen controller in addition to its analogue and digital sticks, pads and buttons; and recently released *Steam Deck* from Valve Corporation (2022), which also features analogue sticks and touchscreen [383]. Nintendo Switch and Steam Deck can be considered both handheld and stationary consoles, as they can be played as portable devices or connected to the TV or a PC monitor and act as normal video game consoles. Nowadays, even though there is a vast choice of touchscreen-based smartphone games, handheld devices are still popular in the video game industry.

Gaming and aviation (particularly flight simulation) industries have been intertwining for more than a decade, fuelling the rapid rise of the computational power of GPUs[2] and CPUs[3] [338] and benefiting from that by providing better and better visual systems, realistic aircraft models and hardware (plug-and-play joysticks, yokes, throttle levers and rudder pedals) to control them. The market convergence can be seen in pilot training schools using commercial off-the-shelf (COTS) software such as *FlightGear Flight Simulator*, *X-Plane* or *Microsoft Flight Simulator*, with professional-looking hardware such as *Thrustmaster Hands On Throttle-And-Stick (HOTAS) WARTHOG^{TM} Joystick and Throttle*, officially licensed US Air Force A-10C aircraft replica [351], and even more basic like *Logitech Extreme 3D Pro Joystick* [229], which is used in initial, low-fidelity aptitude exercises [4].

---

[2] Graphics processing unit.

[3] Central processing unit.

[4] as seen at the stand of The Honourable Company of Air Pilots, Royal Aeronautical Society event, London, November 2017 [302]

2.3.2 Rotorcraft

Contrary to fixed-wing aircraft inceptors, rotorcraft control was not "standardised" from its beginning – due to additional controls (cyclic for lateral and longitudinal control, collective for vertical control and anti-torque pedals for yaw control), more aspects were needed to be considered. The concept of vertical lift generated by rotating "blades" can be traced back to a Chinese toy invented about 400 years BC. The idea of a stick with attached feathers, which could "fly" when spun between hands, was most probably inspired by the seeds of a sycamore tree. A similar design was also found on da Vinci's *aerial screw* in the 15th century. 4 centuries later, a few different designs came to light, such as Cayley's *Aerial Carriage*, Ponton d'Amécourt's *helicopter* (first noted use of that term) or Edison's machine. None of them, however, enabled a human to fly. This was achieved by the Breguet brothers with *Gyroplane No. 1* (1907), not long after the first successful flight of the Wright brothers' plane. Their design was very unstable and required four men standing on the ground to hold it steady. Throughout the years, helicopters were gaining more features that enabled them a steady flight, such as more rotors, a gyroscope, and autogyro implementations, along with von Baumhauer's inventions - the cyclic and collective control [210, 373]. Each of them had a different set of controls, though. In 1940, Sikorsky developed the *VS-300*, which set the standards for conventional rotorcraft controls [147]. Since then, helicopters have been steered using the same set of inceptors: a control stick for cyclic control, a collective lever for vertical control and pedals for yaw control. In recent years, fly-by-wire technology enabled researchers to carry out a series of studies which focused on investigating novel steering concepts for rotorcraft vehicles. One of the ideas was to replace a conventional stick in rotorcraft with a steering wheel. The results have shown that there is a possibility of reduction of workload and increase of handling qualities, especially within inexperienced participant groups [147, 186, 265, 317].

2.3.3 Land transport

Similarly to aircraft's yoke/sidestick and rudder pedals, the car's control method concept has not changed almost since the beginning of the automotive industry's history. Although the early automobiles, with Benz *Patent-Motorwagen* (1885) being the first of them [248], used a tiller to control them, the emerging automotive industry quickly adapted the idea of the steering wheel. This was first introduced in the *Panhard et Levassor* car in 1894 and almost completely replaced the tiller by 1908 [131]. Gear, brake and throttle pedals were not standardised in the beginning. In some vehicles, pedals acted like a gear transmission system, and the throttle was controlled by a hand lever mounted to the steering wheel. One of the first cars to have a "modern" pedal arrangement was Cadillac Type 53 (1916) [191]. Since then, the majority of changes, upgrades and research included "under the hood" changes, not the inceptors themselves – a situation similar to aircraft's inceptors.

The idea to change the conventional steering wheel and pedals into something else was already explored over 18 years ago by Haas & Kunze [148], Andonian et al. [13], and Kelber et al. [188], but until recently has not seen much more research. In the last few years, its potential was evaluated in the works of Saupp et al., who tried using active "drivesticks" in a car [313, 314], Large et al., with their investigation of using joysticks as an alternative to traditional control [207], or Lindner & Greul, who made more general assessment on steering wheel's future [225]. There was also a study about completely removing the steering wheel (Schaefer & Straub [316]). Industry representatives made efforts in that field as well: *Dagne*, a joystick-steerable, electric vehicle concept developed by Werner and Sandoz from Multimode Technologies that won the IDTechEx future of electric vehicles award [33]; and *Sayer*, a new steering "wheel" from Jaguar, presented during Tech Fest at Central St Martins, University of the Arts London [177]; and even in the previous century, in the 1990s, Saab has built a concept car - Saab *Prometheus*, steerable by a single joystick [133]. Recently, Mercedes proposed replacing the steering wheel with four joysticks in their Mercedes Benz F 200 *Imagination* [178]. Even though not all of those concepts were successful, there are more and more attempts to change that, owing to the technology moving forward. The main reason for it is that with the constant rise of automotive technology, especially Drive-by-Wire and autonomous or self-autonomous driving, driver's comfort and ergonomics might be significantly improved. Greathouse presented an overview and raised interesting points on the topic, along with the graphical representation in Fig. 2.4 [145].

Contrary to flight simulators, driving simulators are not as old as their respective industry itself. This is mostly due to the fact that it is much safer to test a new car than a new aeroplane or helicopter. When a car breaks down, it is still on the ground, and there is much less chance of a fatal accident. First car simulators were developed in the 1930s, and with their evolution, throughout the years, they have been serving

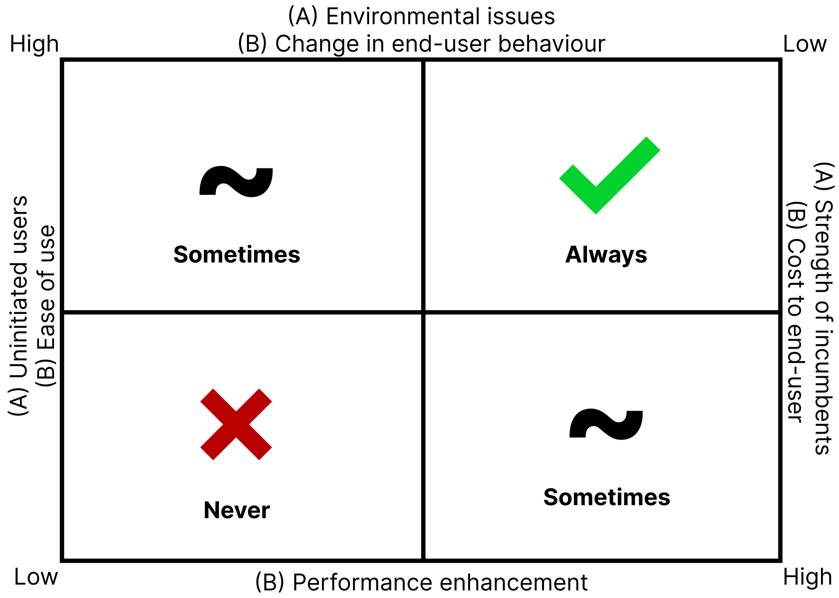


Figure 2.4: Environmental issues (A) and product factors (B) to consider when trying to reinvent the wheel. Adapted from Greathouse [145].

as a test-bed for a variety of research, such as driver behaviours, variable message signs, in-vehicle systems, and automated vehicles [393]. Bouchner gives a detailed insight into driving simulators in their journal article [52]. A systematic review done by Wynne et al. has proven that modern driving simulators lack the validity and fidelity standards [393], which, on the contrary, are present (and very strict) in civil flight simulators [118]. Cars are becoming increasingly sophisticated in technology and functionality, with many modern vehicles featuring advanced driver assistance systems and connectivity capabilities. This has led to the car becoming a more centralised hub for various operations, similar to the evolution of aircraft cockpits in the past. Researchers and industry professionals are now exploring new ways of controlling these vehicles, including the potential replacement of traditional steering wheels with more compact and ergonomic alternatives, in order to improve the driving experience. This interest in new control methods extends to other types of vehicles as well, including aircraft. Understanding the implications and potential benefits of these new control systems is important for the continued development of these technologies. An example of such alternative approaches in land transport is a concept of tangible interaction, with an AI[5]-controlled interface that allows for tangible interaction in car systems'

---

[5] Artificial intelligence

controls [135].

2.3.4 Maritime transport

Maritime (water) transport exists along with humanity's timeline. As archaeological evidence suggests, from the first simple rafts built by Homo erectus more than a million years ago [366], through the oldest known boat, the *Pesse canoe*, dated to be almost 10 000 years old [349] to gigantic cruise and warships of today, it played a major role in the history of mankind [276]. The first controllable watercraft did not use a steering wheel – instead, they relied on a tiller or whipstaff, which were levers attached to the rudder stock. This was until the 18th century when it is believed by historians that the first steering wheel was implemented. It was connected to the rudder by a system consisting of an axle, pulleys, and ropes. It quickly became a standard, although their look, position on deck, and underlying systems have evolved over the years [301]. Similarly to control systems in aviation and land transport, one concept defined how watercraft is controlled and is in use up to today. To the author's best knowledge, there is a lack of research into alternatives to the ship's wheel proposed to date.

2.3.5 Mobile phones

First mobile phones did not need much interaction from a user – their only purpose was that of the landline telephones, which was simply to make a call. Early mobile phones history situates between Tigerstedt's first patent for a "pocket-size folding telephone with a very thin carbon microphone" in 1917 [322] and the first commercially available cellular mobile phone, Motorola DynaTAC 8000X (weighting almost 1 kg), entering the market in 1983 [269]. In the first handheld phones of the 1980s and 1990s, a "controller" which enabled human-machine interaction was just a numeric dial pad with several additional buttons that allowed the user to confirm or cancel a call. Soon after that, the miniaturisation process began (and lasted until all-touchscreen phones started to gain popularity; then, the trend inverted) [201, 385]. In the meantime, text messaging, mainly in the form of a short messaging service (SMS), was developed. The first SMS message was sent in 1992, and within the next decade, it became a standard for all mobile phones to include this service [328]. In order to write a message, the user needed letters, not just digits. Fortunately, the E.161-standard type keypad was present in phones as early as in the late 1980s and early 1990s (Nokia Mobira

Cityman 900, Samsung SH-100, Motorola MicroTAC 9800X), taking over the idea from 1930s rotary dials with three to four letters assigned to each ”2” to ”9” digit [174]. The first mobile phone to enable writing messages easily is believed to be Nokia 2010, released in 1994. To shorten the time required to write a message, various predictive text systems were developed, mainly Tegic Communications’ T9 in the late 1990s and Motorola’s *iTap* in the early 2000s [94, 268]. Nokia Communicator 9000 (1996) was the first mobile phone to feature a QWERTY keyboard [28]; however, this keyboard layout started to gain popularity in the 2000s, owing to BlackBerry and Nokia E61 and its succeeders. They are considered the first ”smartphones”, as they could offer more features than just calling and texting, such as push e-mail, web browsing, and multimedia functionality [290]. This expanded the need for human-machine interaction to be more complex. Over the years, touchscreen technology has superseded the physical input methods in mobile phones. The change was not sudden, though – there were also hybrids of QWERTY and the touchscreen like the BlackBerry Q10 or HTC Dream (the first phone to feature Android as an operating system) [201]. The device that is, however, considered the first ”true smartphone” was IBM Simon, the first handheld device with phone features and touchscreen control, which entered the market in 1994 [307]. Nevertheless, touchscreens needed another 13 years to take over the mobile technology industry.

The iPhone, released by Apple in 2007, was the first successful commercial product to popularise touchscreen technology, marking the beginning of a new era. Prior to this, touchscreens had been mostly used in specialised applications such as ATMs and retail kiosks but not in portable devices. Apple’s introduction of a touchscreen into the smartphone market forever changed how users interacted with their devices. This technology, combined with the intuitive iOS interface and other features, quickly made the iPhone a massive success and popularised touchscreen technology for the masses [20, 290]. With every major company wanting to earn their shares on the market and with advancements in touchscreen technology, described in Section (2.5), touchscreens quickly became almost the sole control method in mobile phones. In 2013, global smartphone sales overtook non-touchscreen phones (”feature phones”) [359], and in 2021, they had a gigantic advantage (with revenue of 0.48 trillion USD and volume over 1.5 billion pieces of smartphones versus revenue of 11.57 billion USD and volume 284 million pieces of feature phones [332, 333]).

For over two decades, numerous studies have been carried out in the field of human interaction with mobile phones. Silfverberg et al. compared text input speeds after the introduction of the T9 system [326]; Kiljander investigated how previous experience with mobile phones affects the usability with a new mobile phone interaction style [190]; Finley compared different layout styles on smartphones in terms of usability and proposed website formatting for phone's screen size [126]; Rajput et al. analysed the design of smartphone interfaces in terms of human-computer interaction [288]; Moreover, there has been some research on improving the accessibility of smartphone interfaces for disabled persons [102, 175, 240, 253, 320], as well as elderly [21, 308].

2.3.6 Industrial, agriculture and medical industry

Joysticks are widely used in operating industrial machines such as assembly lines, cranes, or excavators. Burgess-Limerick et al. conducted an experiment in which they compared the steering wheel and joystick in a shuttle car used in underground coal mines [56]. Joysticks are present in most heavy machinery control panels (although nowadays, they are progressively being replaced with touchscreens). Construction and agricultural machinery require a lot of learning. An example of its complexity can be seen in the first episode of the documentary "Clarkson's Farm", where Jeremy Clarkson, long-time presenter of popular TV shows "Top Gear" and "The Grand Tour", operates a tractor for the first time in his life. In the episode, he points at all the joysticks, buttons, knobs and levers in the tractor's cockpit, commenting that he does not know what any of them do, finally giving up and calling an expert for help [75]. This situation raises a question: *how can one make it easier and more understandable?* Such a question has already been addressed in research with regard to touchscreens, for example, in Kivila's work, where the author compared conventional and novel touchscreen crane control. Results have shown the better performance of the latter [192].

Usability of control is also an important factor in the design of medical equipment. Interestingly, there has been a focus on the user-centred design approach in designing medical devices [339]. Lewis et al. have shown that users are adapting to new medical systems faster by using touchscreen interfaces [213]. This observation can be applied in different areas as well. Joysticks are also used in electric-powered wheelchairs, and modern wheelchairs can even be operated using a touchscreen [101].

2.3.7 Summary

The presented literature review shows that nowadays, many devices are controlled either by a joystick or touchscreen. In the video game industry, especially for game consoles like Xbox and PlayStation, gamepads are the leading inceptors. Gamepads, specifically as video game controllers, have been a study subject for a long time [6, 142, 249, 263]. However, the gaming industry is not the only field where gamepads are used. They have been adapted in other areas, such as the military (gamepad-style controllers are used for drone control [104]; Xbox gamepads replaced ”heavy and clunky” helicopter-style joysticks to control periscopes on The United States Navy submarines [227, 243]), robotics and industry (as a device to control an industrial robot [374] or even a team of robots [49]), teaching (with a gamepad as a physical computing device [45]), biotechnology (as a control for molecular visualisation software [166, 258]), or medicine (vibrotactile gamepad was used as a therapeutic aid for autistic children [67]; cognitive-sensory-movement gamepad used in therapy for children with ADHD [237]). Adapting such off-the-shelf equipment is money-saving and easy to use. The gaming industry is vastly popular, so many people are already familiar with this type of controller. Moreover, its usability and learnability are very high even for people who have never used it, which is evident from the results of this study (presented in Section 5.7.15).

Another reason for the choice to use this type of controller in this study was that some research has shown that gamepad and touchscreen usability and performance can be similar [271], or that gamepad can be better [49, 273, 274], although users have reported different experiences while using each of the controllers [142]. Apart from gamepad-touchscreen comparison studies, there were a number of publications comparing a joystick with a gamepad as well [258, 303, 304].

Thus, this study investigates aircraft control using: a joystick (in this case: sidestick native to the flight simulator used) as a conventional control method; a gamepad, as this is the most common analogue controller in video gaming; and a touchscreen as a novelty and radically different proposition. Kiljander made an interesting observation in the epilogue of his doctoral thesis, in which he investigated the mobile phone interaction style evolution and usability – the author was astonished at how fast his nine-year-old son changed from one mobile phone to another with the different interface (interaction logic); and it happened back in 2004 or earlier, where the smartphones were not as common as they are today. The author said that this ”transfer from the old interaction style to the new one had been completely natural and seamless" [190]. This shows the importance of learnability and usability in any system, whether it is a mobile phone, video game, heavy machinery, car, ship or aircraft control method, as well as the fact that "digital natives" are getting used to touchscreen-operated devices from a very early point in their life, thus being much more optimistic towards that technology in various areas, not only mobile phones [62].

2.4 Evolution of the flight deck inceptor

Aircraft control systems have evolved exponentially within the last few decades. The history of aircraft inceptors begins with the Wright brothers and their first aircraft prototypes. The basic idea of aircraft control principles has not really changed since their invention; it was just the technology to make it easier that was developed. In 1903, the Wright brothers designed and flew the very first aeroplane [391]. In it, they were using a pulley system to control it. It used warping wings instead of ailerons, the elevator was on the front, and it had two single-axis sticks to control the rudder and elevator plus a hip cradle, where a pilot was laying down on their belly to control the warping of the wings. Compared to the next generation of planes, the only differences in handling were the roll (instead of a hip cradle, the rudder pedals and combined 2-axes column for roll/pitch were introduced) and the roll movement, which was done by ailerons instead of wing warping. One of the first aircraft to introduce the control by double-axis column was Avro 504 (1914-1932) – the pilot was in a seated position, and it had rudder pedals, the solution known up to today. However, it still used a relatively simple pulley system, so controlling bigger aeroplanes was not possible because of pilots' fatigue after short flights.

In the 1940s, Avro Lancaster Bomber was produced, introducing hydro-mechanical flight controls to change the position of flaps; however, the inceptors (column/yoke and rudder pedals) idea stayed intact. Between 1945 and 1960, with the further development of hydro-mechanical controls, it was possible to build bigger aircraft, like Boeing 707 (1958), which used hydraulic power to control all three control surfaces. The concept of inceptors still has not changed, though. Over ten years later, in 1969, Boeing 747 was the first aeroplane with hybrid hydro-mechanical systems (with a fully powered and functioning actuation system [2]. The years 1970-1994 brought the introduction of a fly-by-wire (FBW) system, which was an electronic system that replaced manual flight control. The wired, electronic system controlled the hydraulic systems. Mirage 2000 (1978) is one of the first aircraft to feature a FBW stick as a control system [167]. The main difference with the previous architecture was that much less force was needed to operate the stick.

Airbus A320 (1987) was the first airliner with FBW and sidestick controls, and it is still in use today. Since 1994, there has been much research into new flying qualities and HQ technologies, such as the use of fibre-optics instead of wires [196] and touchscreen controls – for example, in Gulfstream G500 (2018) [180]. Nowadays, it is hard to determine the percentage of aircraft that have a sidestick instead of a yoke. Each control system has its own advantages and disadvantages, and the choice usually depends on the aircraft class [122]. Light-weight propeller aircraft tend to have a traditional central column, sidestick can be seen in business and fighter jets, while the airliner category is mostly divided between Boeing, which uses a central yoke, and Airbus, which uses a sidestick [199].

Conventional sidestick has been in use for over 30 years, and much has changed "under the hood". Aircraft control systems have evolved exponentially within the last few years, even though there was the same basic principle, to put it simply: "there is a stick or column. When pushed, the aircraft pitches down. When pulled, the aircraft pitches up. When moved to the left or right, the aircraft will roll in that direction". The so-called control surfaces manipulate the lift and drag in different directions, giving control over the aircraft [204]; however, there is none-to-little research on the sidestick itself. The touchscreen controller, as well as studying the effect of using a console gamepad, are the focus of this work and contribution towards novel research in the inceptor design discipline. The idea of having a joystick to steer aerial vehicles is so deeply rooted in the human mind that even some of the futuristic pop-culture productions depict futuristic space shuttles that still use joysticks as a steering method. An example can be seen in a space opera TV series, "The Expanse". The story is based in the 24th century, with technological advancements that let people travel between planets in the solar system within minutes [1]. And yet, sidesticks are a common sight throughout the episodes, despite the fact that large, advanced touchscreen panels, gestures and voice commands are widely used in that universe.

2.5 Touchscreen technology

The history of touchscreen technology dates back to the mid-twentieth century and, similarly to modern simulation, also started with aviation. The first technology to enable the user to interact with a computer without a physical joystick, knob, keyboard, or mouse was developed in 1965 by E.A. Johnson. His invention, called the capacitive touchscreen, used an electrostatic field to detect the presence of a finger on a panel. This was made for air traffic control operators in an attempt to improve the decision-making process and reduce workload [183, 184, 272].

In the 1970s, several companies developed touchscreen technology, but it was not until the early 1980s that the first commercial products were released. In 1983, Hewlett-Packard released the first widely-available computer with a touchscreen - HP-150 [169]. The first touchscreen-based application was a *ViewTouch®*, a graphical touchscreen point-of-sale software, which ran on Atari ST computer with a capacitive MicroTouch touchscreen [259]. This was the first step to self-service kiosks and ATMs that are common nowadays. The software was used to make orders directly on a computer screen, allowing customers to select items and enter payments without needing a cashier. Since the 2000s, touchscreen technology has paved its way into a wide range of devices, including laptops, tablets, smartphones, and even televisions. The popularity of the iPhone in the late 2000s has had a significant impact on the way touchscreens are perceived. Using finger gestures to interact with a device has become the accepted norm, with many other companies adopting similar designs. This has not only been felt in the consumer market. The touchscreen technology advancements have also had an impact on the scientific and academic communities. In recent years, there has been a push to develop touchscreen interfaces that are more intuitive and natural, such as gesture-based interactions [70]. This research has been driven by the desire to create more efficient interfaces that can be used in a variety of applications. An example can be found in the aviation industry, where touchscreens are making a ”full circle” after their first conceptualisation as an air traffic control interface in the 1960s, as presented in the subsequent section.

2.5.1 Touchscreen in aviation

Touchscreens are more popular than ever, owing to smartphones and tablets [378]. This means that new generations will be familiar with this technology since their childhoods. So far, touchscreen interfaces have been proven to be effective in various areas, such as communication, transport, medicine, and the heavy-duty industry. They have not been explored much in aviation, although some initial attempts were made as early as in the 1980s, not long after the introduction of glass cockpits in F/A-18 Hornet and AV-8B Harrier aircraft [251]. Lately, flight deck touchscreen human-machine interface (HMI) has been introduced in single cases like Gulfstream G500 and G600 [378], and Airbus A350 [123]. Touchscreens also found their way into the flight deck through the use of iPads as ”electronic flight bags” [337]. Harris described touchscreen technology as one of the emerging capabilities that can be implemented in the ”Future Flight Deck” [153]. Chen et al. described touch interaction as a ”Natural User Interface” (NUI). They have pointed out that NUIs showed a high potential for positive research outcomes in novel human-computer interaction methods [70]. One of the advantages is that it can be adapted to each user individually, according to their requirements and preferences, as opposed to physical haptic inceptors [82]. The ”Open Flight Deck” project [46] also explored the potentiality of touchscreens in the cockpit, as well as the ”Advanced Cockpit for the Reduction of Stress and Workload” (ACROSS) project [260]. Studies have shown the relationship between the ”clutter” on a cockpit display and the pilot’s experience [187]. Touchscreen displays offer a way to ”de-clutter” the cockpit, showing only necessary information upon the pilot’s prompt and within their direct input gaze. Touchscreens can offer an increase in performance and a decrease in workload, which has been proven on multiple occasions [115, 300, 331], but, on the contrary, can also cause the opposite in certain conditions [219]. There is a lot more research about touchscreens in aviation, for example, exploring tactility in the cockpit [211, 212]. One of the propositions is a ”foldable”, accordion-shaped touchscreen panel prototype that provides tactile feedback, often requested by the pilots when assessing touchscreen’s usability [58, 280].

2.5.2 Safety and stabilisation

> *Stability. The inherent quality of an airplane to correct for conditions that may disturb its equilibrium, and to return or to continue on the original flight path. It is primarily an airplane design characteristic.*
>
> – *Pilot’s Handbook of Aeronautical Knowledge* [122]

In today’s aviation industry, touchscreens in the cockpit are becoming increasingly prevalent. From navigation to communication systems, touchscreens are being used in various ways to make the cockpit more efficient and user-friendly. However, with these new technologies come new safety and stability concerns. This section explores the ways those aspects are addressed, from the design process to the implementation of the technology itself.

The safety of aircraft is of paramount importance, as they are "safety-critical systems". Therefore, any new technology or system installed in an aircraft must be thoroughly tested to ensure that it meets the safety requirements of the aviation industry. This means that there are rigorous requirements and certification standards that need to be met when designing or modifying an aircraft or any of its complex systems. Although the touchscreens in various forms have already been approved for commercial use by EASA and FAA [17, 81, 123, 396], there are still concerns about how they affect the safety and stability of the aircraft. A multitude of research regarding touchscreens in the flight deck focused on those aspects. Dodd et al. investigated the touch screen target positioning impact on pilot performance in turbulent environments [106]. Coutts et al. assessed the use of touchscreens with turbulence in a 6-DOF flight simulator, where they showed that the usability and workload when using touchscreen input were comparable to other input methods, such as a trackball or fixed touchpad. The study also recommended avoiding elements like sliders and only using short single-touch elements like buttons or switches. The study has highlighted that, as a result of the introduction of glass cockpits, there is a huge potential to replace many physical switches, gauges, and indicators. This will allow rapid modifications and updates by manufacturers. Moreover, touchscreens can reduce head-down scan time, increasing situation awareness (SA) and safety [87]. Alapetite et al. tested the stability aspect by mounting a touchscreen monitor on a rollercoaster ride [7]. A study by Baldus & Patterson has demonstrated that there is no significant difference between a mouse and a touchscreen controller in task performance in a vibration environment [32]. Cockburn et al. studied the safety and stabilisation of touchscreen interactions in turbulent situations and analysed different stabilisation techniques [77]. After experimenting with bezel edge grip [78], they proposed "braced touch", which allowed the pilot to rest four fingers directly on the screen and to interact with the fifth finger (either by double tapping, holding or force tapping) [277]. Their trials confirmed that this type of interaction increases performance in high-vibration scenarios. Furthermore, they have proven that using a "lift-off" contact action is most accurate in this environment, earlier pointed out by Potter et al. [285], and Ren & Moriya [291]. The "Symmetry

Flight Deck", incorporated in Gulfstream G500 and G600, features touchscreen panels. Its designers were aware of concerns about the touchscreen's usability in turbulent conditions. To address this, they carried out a series of tests in a van on a bumpy road, as well as in a motion flight simulator and airborne, and concluded that bevels and plinths provided sufficient grip to achieve hand stability and precision. Furthermore, Gulfstream's touchscreens are resistive-type (as opposed to capacitive, which are most common in smartphones). This reduces the chances of accidental interactions, as it requires applying a little pressure to the screen [352, 378]. Finally, there are some concerns about glare, which is created by sunlight and can negatively affect the view of the cockpit instruments [236, 295]. Those issues were already addressed in later glass cockpit studies [378].

In summary, there are numerous papers mentioned in this section which address the problem of safety and stability of touchscreen technology in aircraft. Therefore, in this study, it was decided to omit the limitations of precision caused by touchscreen panel vibrations and focus only on the control experience.

2.6 Urban air mobility

The emergence of urban air mobility (UAM) has the potential to revolutionise transportation, allowing people and goods to travel quickly over congested roads and through urban areas [34, 55, 138, 297]. But with this new technology come some challenges, particularly regarding the safety of these new forms of transport [228, 267]. One of the critical elements of UAM is the use of electric vertical take-off and landing (eVTOL) aircraft, capable of taking off and landing vertically and flying autonomously [74]. Implementing these new aircraft means it must be ensured that they can be operated safely and with minimal disruption to existing airspace. This is where control systems come in. They are used to control the aircraft and ensure that they are used safely. These systems can be divided into two categories: onboard and off-board.

Onboard control systems are those that are installed on the aircraft itself. These systems are responsible for controlling the aircraft's navigation, propulsion, and other functions. Examples of onboard control systems include flight control computers, autopilots, and navigation systems. These systems are designed to allow the aircraft to be operated safely, even in the absence of a pilot [53, 206, 209, 363].

Off-board control systems are located outside of the aircraft and are responsible for remote control. These systems are often connected to a ground-based centre, responsible for coordinating the aircraft's movements and ensuring that it is used safely. Examples of off-board control systems include air traffic control systems, air traffic management systems, and command and control systems. These systems ensure that the aircraft is operated safely, both in terms of spatial navigation and in avoiding collisions with other aircraft or objects [69, 129, 356].

In addition to the two types of manual control systems, UAM aircraft will also require an autopilot system. This system controls the aircraft's navigation and propulsion and is designed to operate autonomously based on the input parameters. Autopilot systems are designed to be highly reliable and accurate, as they are responsible for ensuring the safety of the aircraft during the cruise. Through the use of these control systems, UAM aircraft can be operated safely and with minimal disruption to existing airspace [53, 74, 206].

Some previous tests have shown that the pilots had to use a significant amount of strength when changing the altitude using a sidestick [107]. A gamepad, or another controller with similar design philosophy, could ease that problem, as the thumb/finger sticks provide much less friction. Alternatively, a touchscreen controller in a more comfortable position could also cancel the fatigue in this case. There is evidence suggesting that there is a need for specific alternative controllers for unmanned aircraft. So far, studies have not explored touchscreen options; there were only investigations of passive and active sticks, mouse-like devices, and gampepads [129, 306]. The UAM market is increasing with new battery and electric vehicle technologies being developed [143, 298]. There is an increasing demand to develop new flight control systems that are low-workload and easy to learn as most eVTOL pilots will not have much experience when UAM eVTOLs become commercially available [231]. Many studies are focusing on hybrid/electric propulsion and other internal aircraft systems [4, 235, 380, 397], or general design [284]. However, there is very little research on hybrid-electric aircraft indication systems. In off-board systems, there is no problem with turbulence; moreover, the touchscreen area can be adjusted, thus making the input more precise.

3

Future Systems Simulator

> P. Szumowski: *What do you do for a living?*
> W. Korek: *I work on flight simulators.*
> P. Szumowski: *Do they tell you to 'stop simulating' when you want to take sick leave?*
>
> – P. Szumowski, comedian[1]

3.1 Motivation

In traditional aircraft, the benefits from optimising airframe and propulsion systems independently are diminishing [63, 109, 208]. The integration of those systems in novel, complex platforms, such as electric vertical takeoff and landing (eVTOL) vehicles, presents a unique challenge [8, 95, 125]. These systems are often developed by separate companies, each with their own intellectual property agendas and constraints on their development activities [111]. For example, using simulators based on existing aircraft or depending on external suppliers limits the possibility of rapidly iterating through flight control and cockpit display concepts. This can lead to suboptimal solutions and a lack of consideration for the integration of the "human-in-the-loop", potentially resulting in a poor user experience or even disastrous consequences [239]. The increasing complexity of automation and a shift in the role of the pilot from aviator to "mission manager" in conventional aircraft further highlights the importance of considering the integration

---

[1] *Enzymy i Pioruny.* Bedford, February 2022

of the human element in the design process [43, 198]. Existing simulators may not be able to adequately address these issues due to their fixed platform architectures and lack of flexibility [369].

Over the years, Cranfield University has worked on integrating flight simulation both in the academic environment as well as with industrial partners [141]. Recently, a new engineering flight simulator (EFS) called the Future Systems Simulator (FSS) was developed, shown in Fig. 3.1. The author was in the team responsible for the design, development and maintenance of this simulator, and his work largely contributed to the outcomes of this thesis. An innovative technique was used in designing the flight deck's visual representation and the physical cockpit fuselage part, which was recognised with an international *iF Design* Award in *User Experience (UX)* category [89]. Moreover, the FSS has received significant interest from top UK flight simulation companies, with representatives describing it as "crisp" and very modern-looking[2]. It is located in Aerospace Integration Research Centre (AIRC) at Cranfield University and it allows for collaborative projects with companies on technical readiness levels (TRL) 6/7[3][119], usually associated with businesses.

The FSS features a highly reconfigurable and modular futuristic cockpit with an all-touchscreen panel that can be rapidly adapted for different research needs. The "default" human-machine interface (HMI) setting was designed to functionally represent a business jet environment based on aerospace standards. The FSS does not only offer "future" technologies to be investigated; it can, for example, mimic Airbus A320 and A350 cockpits in a way that all the operations are familiar to pilots taking part in a study.

This chapter demonstrates how FSS enables state-of-the-art research of various aerospace technologies, from engine systems displays, through the HMI in a flight deck to hybrid/electric aircraft concepts.

---

[2] BAE Systems representative, personal communication, August 2022; Flight Simulators UK representative, personal communication, September 2022

[3] From definition: TRL 6: *technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)*; TRL 7: *system prototype demonstration in operational environment* [160]

Figure 3.1: Future Systems Simulator in a conventional multi-crew configuration setup. Reproduced under the courtesy of DCA Design International [97].

3.2 Background

3.2.1 Engineering flight simulation

Nowadays, aircraft cockpits ceased to be just spaces to control the aircraft's flight path. The complexity of today's flight decks has turned pilots into systems managers instead of just simply aircraft controllers. They need to aviate, navigate, communicate, and administrate - and that list is not exhaustive [43]. Air traffic control plays an important role in that process; however, lost connection or communication can sometimes happen, often leading to accidents [50, 92, 144, 246, 379]. Because of this, pilots have to be trained to operate the aircraft without any external dependencies, and the overwhelming complexity of the aircraft systems can also lead to a potential human error [390]. With the rising popularity of glass cockpits and interactive touchscreen controls in the flight deck, it is possible that in potentially dangerous situations, pilots could be aided by a flight deck, which would provide them with the right information at the right time. However, due to the complexity and strict regulations of pilot training, there is a huge barrier to the adoption of novel flight deck technologies [312, 324, 329]. The FSS was built with the intent to provide proof of concept to these future decks in a cost-effective way.

The past forty years have proven that simulators are invaluable assets in flight training (both civil and military), safety procedures, aircraft design, and research in the aeronautics field [11]. EFSs can be used to conduct scientific research by providing a safe and cost-effective environment to simulate actual flight conditions. They can be used to study the effects of specific flight manoeuvres on aircraft performance and test new aircraft designs, components, and systems. Moreover, the effects of different environmental conditions on aircraft performance, such as turbulence, wind shear, and icing, can be investigated. With the use of EFSs, researchers can gain valuable insight into the science of flight and develop new technologies that improve aircraft safety and efficiency. Additionally, EFSs can be used to train pilots in new aircraft, new manoeuvres, and new safety procedures, thus helping to ensure the safety of passengers and crew. As such, EFSs can be a valuable tool for the scientific, engineering, and industrial areas. However, the engineering flight simulation discipline has not been recognised enough in aerospace-themed academic fields [11]. Having a state-of-the-art simulator available for students and seeing the amount of research carried out on it just in the last few years proves that it is essential for flight simulation to be included in research programmes, and the FSS brings a great contribution to this need becoming a reality. The FSS contradicts what Oberhauser & Dreyer suggested, that the "Engineering Mock-Up" simulators, although high fidelity, lack the flexibility of desktop simulators [270]. The FSS allows changes in the HMI design to be applied in the later stages of any research without altering the project's time constraints. Moreover, it can fast-track the paper-based concepts into the flight deck in a novel approach. This, in turn, enhances the way aircraft systems are presented to pilots by providing the most meaningful information based on a current situation, thus positively influencing decision-making.

3.2.2 Human factors in design - US Airways Flight 1549 case study

The unprecedented event of US Airways Flight 1549 shows the importance of the pilot's decision-making in an emergency. The aircraft experienced a bird strike that damaged both engines, forcing pilots to make a quick decision whether to divert to a nearby airport or land on the Hudson river. Upon the Hudson river landing, all passengers survived, which was described by the media as ”the miracle on the Hudson”. After that event, Eastwood produced a film entitled *Sully*, based on Captain Sullenberger’s book [112, 335]. The film, although ”dramatised”, has depicted the pilots’ decision-making process accurately, according to the National Transportation Safety Board’s (NTSB) report [264]. The pilot ”eyeballed it” - based on his experience of 42 years of flying, carrying over 1 million passengers, which NTSB noted as one of the factors contributing to the accident’s survivability. But if the pilot had not been as experienced as Sullenberger, the outcome could have been more tragic. The pilots only had about 60 seconds to make the decision - and according to post-accident simulations, going back to the departure airport would have been unsuccessful after only a 35-second delay. It would not be possible to go through all the checklists and do all the calculations needed to check if the distance to the airport was possible to make at the altitude they were on. Had there been better HMI systems, the computer might have aided the pilot in decision-making, rendering the whole situation less risky. Even though this landing was ”lucky”, better engine information would have helped with the decision since there was ACARS[4] information that one of the engines was still ”sub-idle”, meaning it could still have some thrust left, allowing the aeroplane to fly relatively safely to the airport. One of the aims of the FSS, which is to provide better engine information to pilots, covers the NTSB recommendation after the incident to ”work with (...) manufacturers (...) to complete the development of a technology capable of informing pilots about the continuing operational status of an engine. (A-10-62)” [264], as well as recommendations from other studies carried out in this field [23, 279]. During the post-accident hearing, Sullenberger mentions human factors (HF). The initial post-accident simulations were carried out in a way the test pilots knew what was going to happen and had the decisions made for them in advance, which Sullenberger pointed out. In real-life situations, pilots are not prepared for unprecedented scenarios. They need to make a decision extremely quickly, which might not necessarily be a good judgment. Having better engine information on board would make that decision safer [264]. Merriman & Karl also underline the importance of HF in the design of glass cockpits and unmanned aerial vehicle control stations [251]. FSS aims to enhance the way aircraft systems are presented to pilots by providing the most meaningful information based on a current situation, thus positively influencing decision-making.

---

[4] Aircraft Communications Addressing and Reporting System

3.2.3 Human-machine interface in aircraft

Numerous studies have been carried out on HMI design in aviation [99, 215, 221, 223, 224, 321, 340, 350]. The primary motivation behind the flight deck design research is the fact that human errors cause around 65-80% of aviation mishaps [43, 110]. Programmable glass cockpits allow better design of the HMI layout, thus potentially decreasing the number of errors made by human operators. However, misinterpretation of the displayed information in a glass cockpit is also the case and reason for incidents [40]. Hence, it is important for pilots to understand how the information is given to them. The information presented to the pilot can not be merely various systems readings as numerous values, gauges, or indicators, thrown in without much thought on the cockpit monitors, as it would be overwhelming or unclear (”clutter”, as a pilot would call it). This would effectively delay or even worsen the pilot’s decision. ”Data (. . .) is not information. It becomes information only when it is approximately transformed and presented in a way that is meaningful to a person who needs it in a given context” [43]. The design of an HMI in an aircraft’s cockpit must go through a series of extensive research. There are multiple ways of achieving that. Some EFSs offer limited capabilities to modify cockpit displays and interactive elements; FSS offers fully customisable touchscreen panels and a possibility to change its elements ”on the go” according to pilots’ needs, allowing rapid prototyping of any HMI aspect. While some studies were offering advancements in HMI design, the trials have been done using just monitors instead of full cockpit representation [134]. FSS addresses those issues and, with the shroud and visual cues, offers a more immersive experience for the pilots, thus increasing the fidelity of proposed novel HMI concepts. Aircraft’s HMI is constantly evolving. Novelties such as voice commands [22, 42, 220] or touchscreen monitors [25, 71, 87, 106, 134, 197, 217, 218] are being investigated, and even already implemented in aircraft [378]. Both of these aspects still have downsides due to relatively new technologies, though - they are not accurate or reliable enough to be widely featured in flight decks. If future technology allows it, however, forthcoming researchers will have a solid study database as a result of current research in those fields. One of the notable concepts in flight deck design was ODICIS (One DIsplay for a Cockpit Interactive Solution). The project aimed to develop a seamless display cockpit with full touchscreen interaction [41]. Thales continued the idea with ODICIS successor, ”Avionics 2020” [347], which then evolved into FlytX, to be integrated into some of the Airbus Helicopters [348].

3.3 Design of the Future Systems Simulator

The design of the FSS was a complex and challenging task, as it had to meet a wide range of requirements and specifications. The main goal was to provide a state-of-the-art EFS that could be used for aviation training, research, and development purposes. To achieve this, the FSS had to be able to simulate a variety of aircraft systems, environments, and scenarios, as well as offer a high level of realism and fidelity.

In addition to conventional simulator capabilities, the FSS had to be designed to support the assessment of emerging and disruptive cockpit technologies, such as cognitive technology and eVTOL systems. This required the development of advanced software and hardware systems that could simulate these technologies and their interactions with other systems in the cockpit.

A thorough design process was followed to ensure that the FSS fulfilled all of these concept requirements. This involved the identification of the key requirements and specifications, the selection of appropriate technologies and components, the development of detailed design models and prototypes, and the testing and validation of the final design. This section will provide a detailed overview of this design process and the steps taken to ensure the success of the FSS.

3.3.1 The goal

Almost all flight simulators are designed to imitate an existing physical aircraft. Their entire reason for being is to replicate the current flight experience for the given aeroplane. Their typical purpose is to support training. As such, they exist to "optimise" the human - a pilot. The assumption is that the aircraft design is correct and fixed, and the human must be changed or optimised (trained) in order for the human-machine system to have acceptable performance.

The principles underpinning the design of the FSS were quite different. Instead of seeking to adapt the human to fit the fixed design of the cockpit, the aim was to explore how iterative changes to the cockpit design could influence the human-machine system performance. The philosophy of the FSS was defined by three core tenets:

1. The cockpit environment should be as flexible as possible, creating the opportunity to explore the way information is presented and the way the pilot can interact with it.

2. It should be fast to iterate - the time to implement a change should not be a barrier to exploring a new option.
3. The design process should be inclusive - no single stakeholder group should own the development process; instead, all stakeholders (pilots, HF specialists, system architects, propulsion specialists, dynamics specialists, and coders) should be able to articulate their ideas.

3.3.2 The process

A core philosophy running through the project was that the graphical user interface (GUI) and the physical components in the system should have parity. This meant designing both in tandem rather than starting with the physical interface. Careful considerations were made as to which controls should be represented on touchscreens and which required more traditional physical elements. This early consideration of the GUI led to changes in the number of displays, their size, orientation and positioning.

The design and delivery of the physical and digital environment required a wide range of skills. The FSS's flight deck was designed and developed by an integrated team comprising of designers, mechanical engineers, electronic hardware engineers, HF and user experience (UX) specialists, and model makers, alongside specialists in simulation, aircraft dynamics, and aircraft propulsion. It was the harmony of the team that proved to be so successful – each group respecting each other's opinion with the power to go off and research specific topics in detail and bring these back to the wider team for adoption into the project. The physical layout of the cockpit displays and controls was informed by the ergonomic needs of the pilot. While this started on paper, with anthropometric mannequins, a full-sized MDF[5] rig, demonstrated in Fig. 3.2, was built early in the project to test assumptions with pilots and to gain additional feedback. This philosophy of stakeholder engagement continued throughout the project. Multiple interface layout iterations were assessed with test pilots in the physical mock-up. The final position, reach, and size of elements were particularly influenced by this feedback.

It was intended that the simulator's HMI graphics (its GUI) should start from a blank sheet of paper unencumbered by existing conventions. With an environment that was rapidly modified and iterated, there was a risk that it could quickly become disjointed. To combat this, an organic-like "style guide" was developed to create har-

---

[5] medium-density fibreboard

Figure 3.2: Future System Simulator's MDF flight deck prototype. Reproduced under the courtesy of DCA Design International [97].

mony between the different elements of the FSS. An appropriate aesthetic strategy was developed that defined all of the appropriate information, colour schemes and how they should be presented. This style guide has been faithfully followed when new graphic elements have been created.

Another key aspect of creating this harmony involved ensuring that the interface worked at multiple levels of abstraction. The UX designers worked closely with the technical teams and the test pilots to establish the hierarchy of needs for the design. This involved considering higher-order goals alongside the need to understand the status of physical components. Key vignettes, typically safety-critical events, were used to stress test this. This focus on presenting what is important, at the right time, most optimally and efficiently guided the team through the detailed decisions when designing the HMI. The author's expertise played a significant role in this process. He was a technical consultant for the way the flight deck elements were developed. Each graphical element was structured in ways that meant that they could be modified easily. The UX team spent time demonstrating how the graphics were built to the wider team. This helped to remove the mystery behind the creation of the graphics, meaning that traditionally non-creative members of the team could feel comfortable challenging and building upon the work. The graphics were created through a series of workshops. Initially, ideas were brainstormed using paper and online co-collaboration tools and then moved through graphic development tools (including *Marvel*, *Miro* and *Figma*) before being implemented in the games engine Unity on to the FSS for testing. This last part was also the sole role of the author. At each step of the way, the solutions were shared with the wider team and optimised based on their feedback. The collaborative nature of some of the digital tools was critical to this. Rather than waiting for formal review meetings, stakeholders could "check-in" and provide comments and suggestions as the design developed.

The physical design of the simulator was designed to be generic in format rather than being aligned to any one particular aircraft type. It was developed to feel realistic and representative to the pilots undertaking user testing. It was also important that it should have an aesthetic representative of a flight deck of the future. The layout of the flight deck's elements is shown in Fig. 3.3. Following some research studies in a range of flight simulator environments, it became apparent that the simulator needed to be more structurally robust than initially anticipated. In demanding flight situations, test pilots could get very physical with the controls and interfaces. As well as being robust, the FSS was designed to be easy to fix. The team designed the physical form of the simulator to be built from rapid prototyped parts where possible. This meant that parts that were damaged or needed to be upgraded could easily be replaced. Using Rapid Prototype parts also helped to minimise simulator downtime due to their low production lead times. The design process was highly iterative and involved the following steps:

- a brief desk-based HF review of the proposed flight deck console was carried out, setting the pilot seating positions and the various optional primary and secondary control formats, locations and orientations.
- once the fundamental layout had been agreed upon, the design quickly moved to 3D CAD[6] review. This was viewed in VR for different sign-offs along the development path.
- several low-fidelity rigs were produced for localised user testing at a range of review meetings and workshops throughout the development process.
- once the fundamental elements of the design had been defined, a basic ergonomic rig of the proposed modular flight deck console arrangement was constructed. This formed the basis of a HF review of the system's ergonomics and usability. At the review workshop, fundamental changes were made to the setup as assessments were made. These changes were made rapidly, allowing the refined design to be re-assessed as part of the same workshop.

Once the design was agreed upon, the project moved into a manufacturing phase.

---

[6] Computer-aided design.

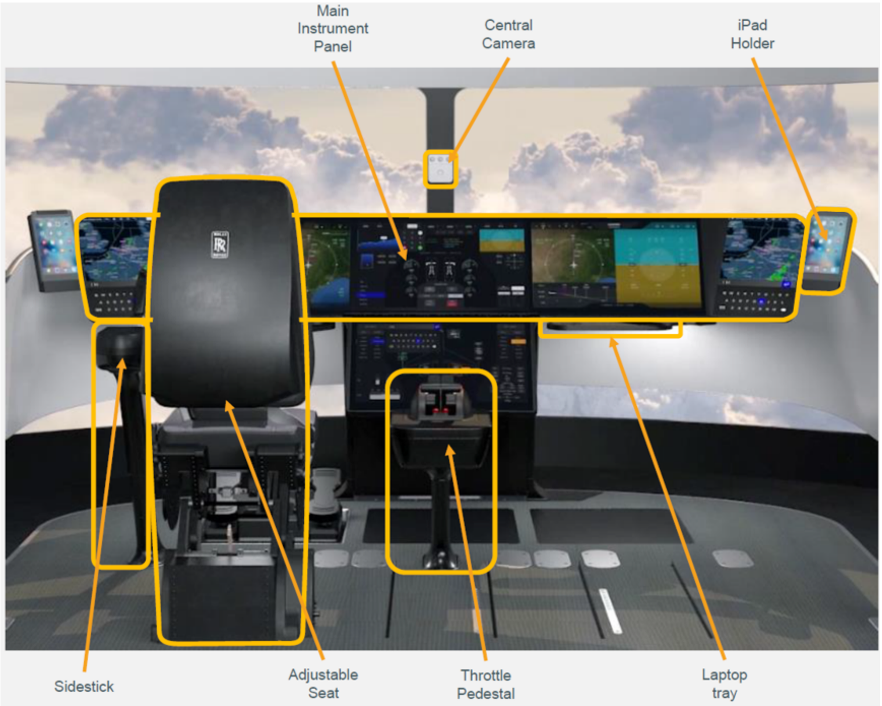


Figure 3.3: Future System Simulator's flight deck elements layout.

This included parts fabrication and assembly. The bespoke parts and sub-assemblies were constructed using a range of low–volume manufacturing techniques. These included 3D printing, CNC[7]-machining, and laser cutting sheet material. Then, the FSS was assembled at the large-scale model-making facility, and once it had been inspected by the key stakeholders, it was transported to Cranfield University's AIRC for final installation, testing and commissioning within the control room space. The author was also involved in this process, as well as in the decisions of the hardware specification used in the simulator.

3.3.3 The future vision

The FSS is a purpose-built "test-bed" designed specifically to explore future HMI cockpit challenges, such as alternative inceptors. From the beginning, the physical and digital configurability was designed to support the broad range of systems in development, from electrifying existing platforms to investigating novel, autonomous eVTOL platforms. It has been designed to be reconfigured rapidly at both physical and digital levels. Adjusting the seating configuration (single, twin, or three side-

---

[7] Computer numerical control

by-side) can accommodate the traditional set of pilots or be used to evaluate future control paradigms, including single-pilot operations with a remote co-pilot. While a touchscreen interface provides the most freedom for future applications, the interface aims to replicate existing cockpit design principles as well as convey a high level of tactile feedback. All critical control interactions were designed to reduce the chance of false inputs. Different physical controls can be added, repositioned or removed. The number of screens can also be changed along with their locations, allowing it to closely represent a variety of civil aircraft control philosophies. The digital elements of the FSS are, of course, highly reconfigurable.

3.3.4 Summary

As part of the global experience of increasing demand for novel products and design methods, the FSS has been devised as a platform that enables concurrent and individual assessment of developing technologies and can adapt to the ever-growing complexity of future aerospace products. Hence, the modularity of the entire simulator framework allows for the integration of any aircraft system as an integral participant of both performance and HF aspects. For this purpose, generic layouts of the synoptic displays of the main aircraft systems were designed, while still adhering to the basic system architecture of the corresponding aircraft category. With such an approach, the modelling of aircraft systems in the FSS can be adapted to the necessities of each simulation scenario, providing the pilots with the required functionalities of the specific system. Moreover, even when a system is not implemented, the displays can be programmed to show dummy data that conform to realistic synoptics.

The cockpit design was developed with the engagement of DCA Design International [98], test pilots and aerospace engineers at every stage of the process to provide the best flying experience while keeping the modularity and reconfigurability of the project. The result is a cockpit made of 6 touch screens, sockets for extra tablets, and extendable trays, as seen in Figure 3.4. The selected seats can be found on in-service Gulfstream G450 aircraft. The seats, sidestick and throttle pedestals can be physically moved to different slots on the flight deck platform. The cockpit "shroud" was designed to provide an immersive feel to the pilots, but it can also be removed if needed.

Figure 3.4: Future Systems Simulator's cockpit close-up render. Every physical or digital element can be repositioned or removed according to research requirements.

3.4 Development and architecture

The FSS development was a multi-faceted and iterative process during which various software and hardware components were designed, prototyped, tested, and validated. The goal was to create a state-of-the-art aircraft systems simulator that could accurately and realistically simulate a wide range of aircraft systems, environments, and scenarios while supporting the assessment of emerging and disruptive cockpit technologies.

To achieve this, a number of technical and logistical challenges had to be overcome, such as the integration of complex software and hardware systems, the ensuring of compatibility with a variety of aircraft models, and the meeting of strict performance and reliability standards. Additionally, the development process had to be flexible and adaptable as new technologies and requirements emerged over time.

This section will describe the key steps and milestones of the FSS development process, highlighting the challenges and successes encountered along the way. An overview of the tools, technologies, and methodologies that were used to design, prototype, test, and validate the FSS will also be provided, as well as the key factors that influenced its development. This will provide a comprehensive understanding of the development process and the considerations that went into creating the advanced and innovative capabilities of the FSS.

3.4.1 Human machine interface and software development

The HMI forming process began with the author's development of several prototypes for the primary flight display (PFD). These designs led to finalising the arrangement of crucial components such as the attitude indicator, navigation display, engine information/gauges, landing gear/flaps digital levers, and other cockpit systems such as the mode control panel. Once the base design was approved, it was implemented in Unity[8] to develop and maintain the HMI easily as this software enables full flexibility in design [252]. Additional features, such as radio control, flight management system, checklists, electronic centralised aircraft monitor, and map, were subsequently added through digitised buttons. The overhead panel, which includes components such as engine ignition and safety switches, electronics, fuel, hydraulic, pressurisation, and other control systems commonly found in aircraft cockpits, was represented as multiple tabs on the central lower display called synoptic pages. The HMI remains in constant development to adapt to the experiment's needs. Moreover, it can represent any existing aircraft cockpit. An example of HMI configuration, based on a business jet's flight deck, is shown in Figure 3.5. There are standards such as CS25-1302/RP-505 for embedded systems in the cockpit [62] and ARINC 661 for the unified preparation of flight cockpit indicators [401]; however, these solutions are limited by existing specifications and aim to speed up the process of preparing virtual cockpits in simulators. The rapid prototyping technique used in the FSS allows for almost non-constrained proposals of novel cockpit elements, such as hybrid-electric indicators.

3.4.2 Aircraft models

The FSS does not use COTS models offered by commercial flight simulators such as *FG*, *X-Plane*, or *Microsoft Flight Simulator*. Instead, it features several flight dynamic models developed at Cranfield University, such as a generic business jet, Airbus A350, or even a novel eVTOL aircraft. Custom models can be more accurate and understandable than frequently "black-boxed" COTS software, which is often hard to validate. The models are compiled, deployed and simulated on the dSPACE SCALEXIO[9], which

---

[8] `unity.com`

[9] dSPACE.com

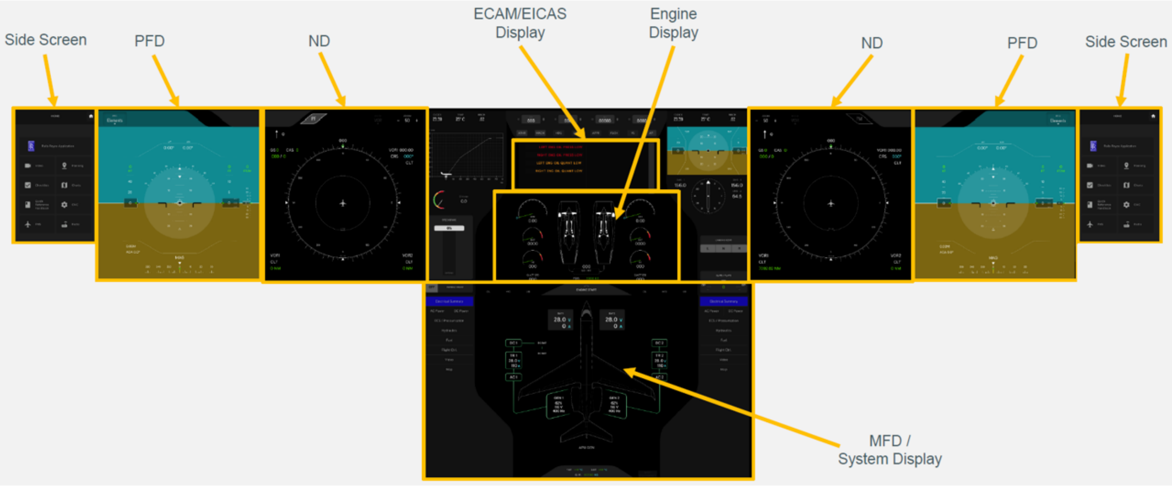


Figure 3.5: Example of developed human-machine interface (HMI) for a business jet aircraft model in the Future Systems Simulator. The flight deck consists of 6 touchscreen monitors, mounted on a stable base. The monitors' layout and the HMI elements can be freely repositioned to accommodate any research requirements. PFD - primary flight display; ND - navigation display; ECAM/EICAS - electronic centralized aircraft monitor / engine indicating and crew alerting system; MFD - multi-functional display.

is a hardware-in-the-loop real-time platform used in a wide range of applications, including aerospace, automotive, industrial, and robotics. The models are created in MATLAB/SIMULINK environment. Although the flight dynamics modelling was out of the scope of this research, one of the author's assignments was to implement a network interface, which would allow communication between the dSpace application and the rest of the simulator. For that, a custom UDP[10] packet had to be defined and coded into the aircraft model and the HMI software. The UDP was chosen for bi-directional communication across the network. Despite its intrinsic flaws, the UDP's advantages include superior speed and efficiency, as this protocol does not need to establish a connection before sending the data.

Through pilot-in-the-loop simulation, meaningful feedback from the end-users of the aircraft system can be integrated directly into engineering development and research. This can therefore enable the prototyping of interface design and control systems early on in the development cycle of a new engine or airframe (or both simultaneously) to allow more informed decision-making during an engineering programme, saving time and costs associated with integration issues that could emerge at much later design

---

[10] User Datagram Protocol

cycles.

3.4.3 Architecture

The FSS undertakes a broad range of computational tasks, which are divided between several computers and connected through a local area network (LAN) using a network switch. This distributed network is depicted in Fig. 3.6. The dSpace unit is configured from the "dSPACE/aircraft model management + Scenery PC", where the different models are compiled to be deployed. This PC is also responsible for generating the outside scenery (using the *FlightGear Flight Simulator* (FG) or *X-Plane* commercial off-the-shelf (COTS) (COTS) flight simulator). Configuration and development of a network interface between the FG and the simulator were also among the author's tasks in this project. The "HMI PC" manages the entire flight deck, which consists of cockpit instruments (indicators), input elements (such as radio, navigation or autopilot panels), and physical inceptors - sidesticks, rudder pedals and engine throttle. The Unity application, created by the author, receives UDP packets from the aircraft model running on dSpace and Instructor Operating Station (IOS), displays the information on the cockpit, and sends pilots' inputs back to the model and IOS. The HMI PC can also stream the cockpit's content to external displays or projectors, for example, for live trial monitoring purposes.

Inceptors

The sidestick controller in the FSS cockpit was based on an off-the-shelf *Thrustmaster HOTAS WARTHOG*^{TM} gaming joystick, but the upper part was removed, and a custom-made handle, designed by DCA Design, was created to ensure the pilot's comfort and ergonomics. The sidestick is also equipped with directional, push-to-talk, and autopilot-disengage buttons. The armrest pedestal for the sidestick is a custom-built unit that is motorised to allow easy adjustment for the pilot's comfort. The rudder pedals are off-the-shelf models from *Logitech's G Saitek PRO Flight* series.

The throttle in the FSS cockpit is fully custom-built, consisting of two independent and motorised thrust levers, a push-to-disengage auto-throttle, and two additional functional buttons. An Arduino processor operates the throttle, which is connected to the PC via a USB port and communicates with it using an RS232 connection. The throttle "communicates" with the rest of the simulator through an interface created by

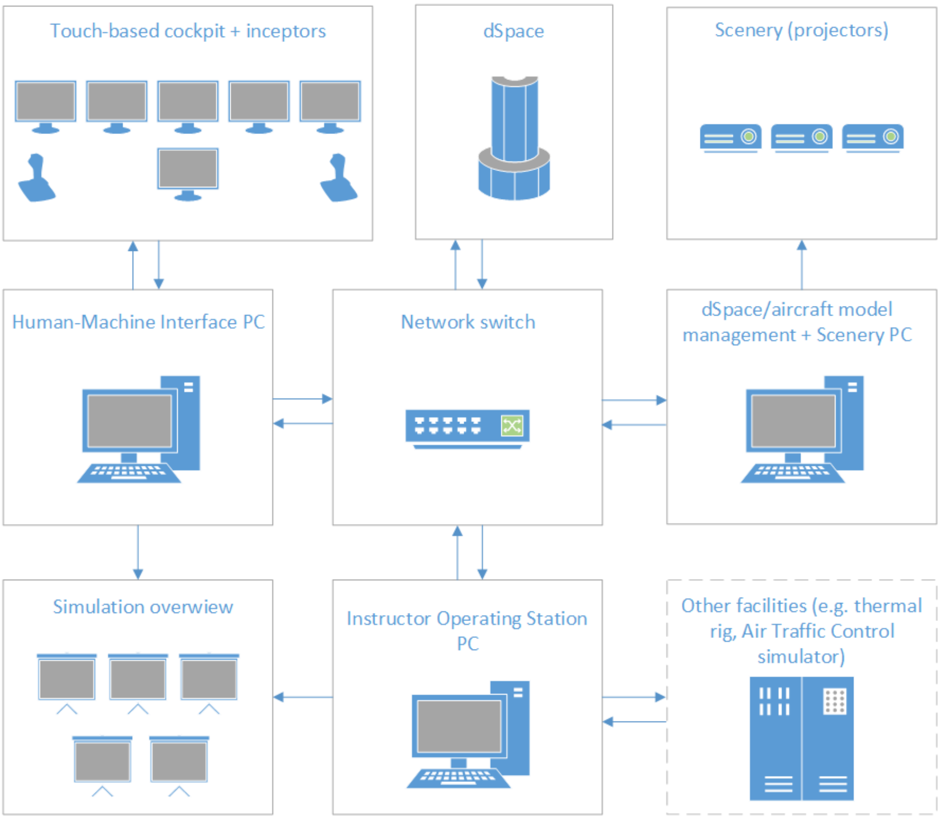


Figure 3.6: FSS's distributed architecture.

the author. Any additional inceptors, such as a landing gear lever or flap/slat control, are integrated within the HMI.

Modularity

The HMI of the FSS is designed to be fully reconfigurable. It was developed using Unity, which offers almost unlimited possibilities in terms of layout and special functions, limited only by the number of touchscreen monitors. This allows for replication of any existing aircraft cockpit for testing purposes, with digitised buttons, levers, and other controls.

One of the key features of the FSS is its portability – it is possible to create a desktop version of the simulator in a short time. The fact that the HMI and IOS software were developed using Unity allows for creating an executable for a specific layout that can be transferred between target computers. The inceptors are off-the-shelf, plug-and-play devices that only need to be remapped in the Unity interface, with the exception of the throttle. However, the software can be quickly reprogrammed to accommodate a regular plug-and-play throttle device. The aircraft models are developed using MATLAB/SIMULINK, which means they can be compiled independently of dSPACE on a regular PC as long as it has sufficient processing power to handle the mathematical complexity of the chosen aircraft model. The simulator can be quickly reconfigured for single-pilot operations or remote-operator setups, as seen in Fig. 3.7 and 3.8.

Figure 3.7: Single-crew configuration.

Figure 3.8: Remote-crew configuration.

Visual System

Data representation is a key aspect of the complete integration of the human-in-the-loop within the modelling and simulation of any technology [338]. The academic and industrial nature of the FSS's vision demanded the creation of an immersive sensorial environment for pilots, in order to become another means to use and validate the data effectively. For this reason, a multi-projector visual display was combined with a cylindrical screen to create an enclosing atmosphere within the cockpit, supported by the flexibility of the FG for image generation.

Currently, the three most commonly used desktop-based COTS flight simulators are *Microsoft Flight Simulator* (MSFS), *X-Plane*, and an open-source FG. FSS required a visual system that was realistic enough to immerse the pilots in their tasks. FG was chosen because of its open-source architecture, vast configuration options and ease of interfacing with the network through a UDP connection. Additional FG capabilities are used to tailor the test requirements to each simulation scenario. For instance, manipulation of weather conditions can be crucial to replicate scenarios for model and data validation. FG provides various means to either control weather at specified regions or to interpret METAR[11] reports around the given airfield. Another feature

---

[11] Meteorological Terminal Air Report

implemented in the FSS was the ability to have other air traffic. Three-dimensional objects can be placed as substitute aircraft at any arbitrary location in the scenery, and the data from the simulator can provide their position information. Ultimately, the networked multiplayer feature will expand the FSS's research capabilities by enabling interaction with other aircraft, flown in other flight simulators (or even sent in real-time from actual in-flight aircraft), in scenarios such as formation flight or air-to-air refuelling.

The visual display system of the FSS consists of three ceiling-mounted Optoma EH515TST projectors and a cylindrical screen provided by 3D Perception[12]. This includes a multichannel image processor and display manager with a user-friendly interface for system setup, control, and maintenance. The screen is a $2.6m$ radius by $2.1m$ height cylinder with a 1.0 gain HD progressive surface to improve edge blending around the curvature. This setup provides a projection with a 200° horizontal field of view (FoV) and a $+21°/-22°$ vertical FoV for full coverage from both Design Eye Points (DEP) inside the cockpit, which can be seen in Fig. 3.9 and 3.10.

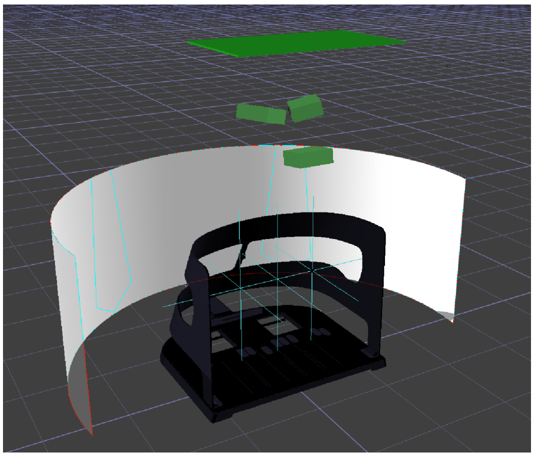


Figure 3.9: 3D view perspective of projectors' setup.

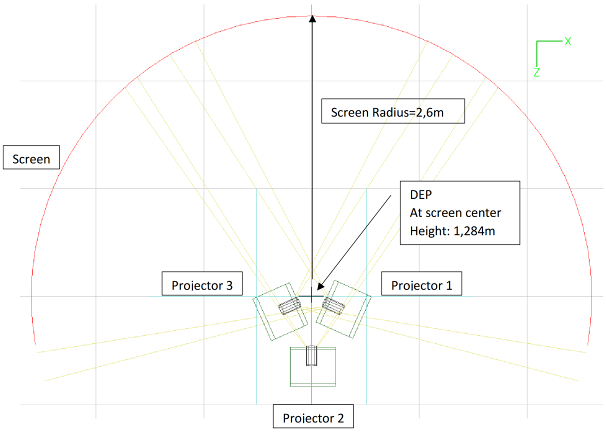


Figure 3.10: System plan view of projectors' setup.

The projection is enhanced by 3D Perception's *nBox* display processor, which receives the video signal from the Image Generator (IG) software and performs re-alignment, colour calibration, warping, and blending onto the screen. The nControl™management software centralises the control and maintenance of the entire display system to ensure consistent high-quality visualisation for any cockpit seating configuration or represented flight scenario.

---

[12] 3d-perception.com

Hardware specification

The hardware specifications for the HMI and visual display systems of the FSS were designed to meet the demands of a high-fidelity, high-frame-rate training environment. The use of state-of-the-art gaming PCs and advanced displays ensured that the FSS can provide a realistic and engaging experience for pilots.

The HMI of the FSS consists of 2D graphics displayed on six 4K LG screens and two HD screens. Four of the 4K screens are part of the cockpit and two are used for the development station. The six touchscreen panels in the cockpit allow for full flexibility and customisation. Four of the screens (two central and two main-side) are 21.5-inch monitors with a resolution of 4096x2304 pixels, and the other two (situated on left and right far sides) are vertically-mounted 13.3-inch monitors with a resolution of 1920x1080. The FSS uses Huawei VR2 cables to transmit video data from DisplayPort outputs on the GPUs to USB-C inputs on the monitors. The PC that runs the HMI is a bespoke system from Renda Solutions[13] designed specifically to meet the FSS requirements. It is water-cooled for quiet operation and consists of:

- Graphic cards: 2x ASUS GeForce RTX 2080 Ti Turbo 11GB GDDR6
- Memory: 4x 8GB DDR4
- Motherboard: Asus ROG Maximus XI Hero Intel Z390 DDR4 ATX
- Processor: Intel Core i9-9900KS 5.2GHz

To generate the outside world imagery, the FSS uses either FG or X-Plane on a state-of-the-art gaming PC equipped with:

- Graphic card: MSI GeForce RTX 3090 Ti Suprim X 24GB GDDR6X
- Memory: G.Skill Trident Z5 Neo 64 GB (2x 32GB) DDR5-6000 CL30 Memory
- Motherboard: Gigabyte X670E AORUS MASTER
- Processor: AMD Ryzen 9 7900X 4.7 GHz 12-Core

The visual display system in a flight simulator is a critical component that plays a significant role in creating a believable and immersive training environment for pilots [12, 59, 353]. To achieve this, the visual display system must have a fast and powerful processor and graphics card that can support a high refresh rate. This helps to reduce latency and improve the sense of immersion by allowing the visual display to update quickly and smoothly in response to the pilot's actions and movements. This is particularly important in situations where the pilot needs to make rapid and precise

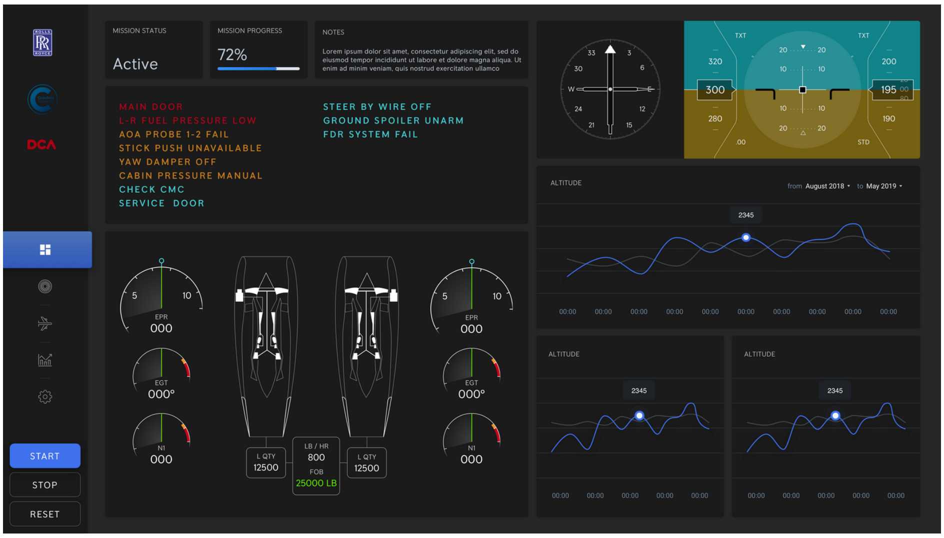
---

[13] www.overclockers.co.uk/renda-solutions

movements, such as during emergency procedures.

3.4.4 Instructor operating station, simulation monitoring and data collection

The IOS allows the operator to control various aspects of the simulation process in the FSS, such as changes in the daytime, weather conditions, other traffic, and flight parameters. The interface is presented in Fig. 3.11. The IOS provides an overview of all aircraft systems and data coming through the simulator. Images from the synoptic pages can be transmitted to a remote control room for monitoring the simulation. The FSS provides multiple views, including HMI displays, video streams from the cockpit (front view of the pilots and rear view of the whole cockpit), live-data graphs, and raw text data.



Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco

Figure 3.11: Instructor Operating Station interface.

The data from the FSS can be live-streamed over the network, allowing observers and researchers around the world to see the simulation in real time. This can be in the form of graphs, plots, flight deck instrument streams, video, or raw data.

All aircraft model and pilot input data are logged on the IOS station at a sample rate of 50Hz. There are two cameras that can record the research trials for supplemental discussion (for example, gesture behaviour analysis, a general posture of the pilot, or verbal feedback). There is also a microphone for audio logging. In order to gain insight into HF and ergonomic aspects of the HMI, an eye tracker can be connected to record the gaze positions of the pilot's eyes.

3.4.5 Classicifation and limitations

The FSS is an EFS that has been designed to provide a high level of fidelity in the research of aircraft systems, environments, and scenarios. Although it was not tested by the European Aviation Safety Agency's (EASA) authority, according to commercial flight simulators' definitions and standards [118], the FSS could potentially be classified as a "flight training device" (FTD) Level 2. This means that it is a simulator that is capable of providing training for specific tasks or manoeuvres, such as takeoff and landing, as well as evaluating the performance of pilots and other crew members. Moreover, the FSS also has many features that could potentially comply with the description of a "full flight simulator" (FFS) Level A and B [118]. For example, the FSS has a high-fidelity visual system, a realistic cockpit layout (albeit represented as touchscreen counterparts), and advanced software and hardware systems that can simulate a wide range of aircraft systems and technologies. This allows the FSS to provide an immersive and comprehensive experience that closely mimics the real-world conditions of flying an aircraft.

The FSS has a fixed base and does not include a motion system. This decision was made based on research showing that motion systems are not necessary for flight simulators to be beneficial in research and training, especially when the main focus is on HF [10, 159, 327]. Motion systems are only necessary when the aircraft's response characteristics are very sensitive and rapid [282], and they can be beneficial for novice pilots training [387]. Additionally, integrating a motion platform with the FSS could introduce problems with cable connections and touchscreen durability, and it would require more maintenance and technical support due to health and safety regulations.

However, the FSS includes limited motion cues in the form of special seat pads that can be attached to the seats to simulate aircraft vibrations (for example, caused by an engine's fan damage). This allows to simulate certain aspects of motion without the need for a full motion system.

3.5 Summary

Over the past century, flight simulation technology has advanced significantly alongside aviation technology. Visual systems have become more realistic, the interaction between pilots and cockpits has become increasingly important, and aircraft models simulated in real-time have become more credible. Previous EFSs were based on existing aircraft environments and systems, with physical levers, switches, and knobs similar to those found in real cockpits. Because of this, these simulators were often physically and financially limited in making major layout changes. The FSS, on the other hand, has been constructed as a highly modular platform that allows for rapid prototyping and testing of radically different concepts in flight deck design, information provision, and flight experience while also focusing on the safety and reliability of all aircraft systems. It can be configured for a variety of purposes, including conventional multi-crew and single-crew operations and unexplored eVTOL-like arrangements. The design of the cockpit of the FSS was guided by principles of clarity of use, precision, and efficiency. To achieve these design goals, a distinctive user-centred design technique was utilised. The use of commercial real-time computer hardware enables expansion for future system integration with hardware-in-the-loop simulation. In particular, future power systems and energy storage concepts could be connected to the FSS through aerospace standard communications protocols, enabling a better understanding of how novel propulsion systems may behave under more representative real-time and flight-dependent load scenarios.

The FSS offers a solution through its flexible, open, collaborative, and human-centred approach to the development of aircraft control and interface design. The FSS's unique capabilities were demonstrated through case studies involving single-pilot operations in a traditional aircraft [137] and the development of an eVTOL pilot interface and control system [103], and further validated using eye-tracking (ET) technology [197, 198, 217–219]. These studies showed that the FSS allows for unconstrained research with early user engagement and the production of scientifically relevant results, which supports the hypothesis (H_2) that **a novel engineering flight simulator helps to streamline the research and validate the results of radically different control methods in an aircraft**.

This page is intentionally left blank.

4

Methodology

4.1 Experimental setup

The experiments conducted in the Future Systems Simulator (FSS) aimed to investigate the differences between the three inceptors considered in this study. To address this, a between-subjects experimental design was employed, with real and naïve[1] pilots recruited as participants. Each participant was tasked with completing four different scenarios in the flight simulator, using three different inceptors: sidestick (SS), gamepad (GP) and touchscreen (TS). The data collected included records of the simulator performance and responses to questionnaires completed by each participant after completing each scenario[2]. The data were analysed using statistical software to examine the differences in performance and subjective experiences between different participants' groups, as well as between the different inceptors. This chapter shows the process of designing and carrying out the author's experiments. It includes the method, participants, environment, inceptors, tasks, trial procedure, and measured factors.

The first ten pilots participated in a "pilot study" [172]. Licensed airline pilots were invited to take part in the trials to gather initial data, but also to refine the experimental setup. Results were presented at the *HCI International 2022* conference and published in the proceedings [197], and further analyses were published in the *Transportation Research Procedia* journal [219].

---

[1] The term "naïve pilots" has the same meaning as "non-pilots".

[2] This research was approved by the *Cranfield University Research Ethics System*, reference number CURES/14853/2021.

4.2 Participants

4.2.1 Sample size

There is no set "optimal" sample size for flight simulation trials. The appropriate sample size will depend on a variety of factors, including the research question being addressed, the type of simulation being used, the experimental design, and the desired level of statistical power. The optimal or minimum sample size will depend on several factors for meaningful statistical analyses, namely analysis of variance (ANOVA). In general, larger sample sizes are typically associated with increased statistical power, which means that the study will have a greater ability to detect a true effect if one exists. However, increasing the sample size also increases the cost and time required to conduct the study. Therefore, the desire for increased power must be balanced with the practical considerations of sample size. There are no definitive rules for determining the appropriate sample size for ANOVA analysis; it depends on the specific research question being addressed, and the characteristics of the data [261]. However, some guidelines have been proposed based on statistical theory and empirical data. For example, one guideline suggested that a minimum sample size of 30 is necessary to ensure that the assumptions of ANOVA are met [80]. Specifically, Cohen recommended a minimum sample size of 30 per group to ensure that the distribution of the residuals is approximately normal, which is one of the assumptions of ANOVA, while Faul et al. indicated that a minimum sample size of 10 per group might be sufficient for small to moderate effect sizes [121].

For this particular study, two options have been considered. The first approach was to involve fewer participants, who would spend more time in the simulator by repeating the trial procedure three or more times. This would ensure the correctness of rating scales such as the Cooper-Harper Rating Scale (CHR) and be less prone to singular accidental mistakes during task execution, which could bias the result. However, with just three repetitions, less skilled participants would still perform poorly. It would take more than three attempts for them to get better, and this would significantly prolong the entire trial process. Similarly, more experienced participants would only improve a little. Furthermore, the longer time spent in the simulator would be exhausting for the participant, which could bias the later results. After discussions with supervisors, this approach was rejected, as the learning curve effect (also known as the "practice effect") would not be steep enough, and the time constraints would not allow for the advised eight or more repetitions [233]. The second option was to have more participants, having only one attempt in each inceptor/scenario combination. The aim was to acquire at least 60 participants with varying flight experience.

4.2.2 Characteristics

The final sample for this study included 74 participants, who ranged in age from 21 to 63 years (mean $M = 31.7$, standard deviation $SD = 10.35$). There were 55 male and 18 female participants[3]. The handedness distribution was: 64 right-handed, 7 left-handed, and 3 ambidextrous (able to use the right and left hand equally well). In terms of flight experience, 27 of the participants held a pilot license (varying from private to airline pilot licenses), were attending a pilot's school or course or were awaiting a rating, with a mean of $M = 2904$ ($SD = 3474$) total flight hours. 38 of the participants reported having at least 10 hours of experience with flight simulators (engineering or commercial, larger than desktop/PC-type), and 38 reported having a positive attitude towards the use of touchscreens in the flight deck (answered ”4” or ”5” in the question ”What is your view on touchscreen technology being introduced in aircraft cockpits? Use the scale from 1 (I do not like the idea) to 5 (I like the idea)”; the overall mean was $M = 3.36$ ($SD = 1.32$). In terms of video game habits, 35 participants reported playing (or used to play) video games at least 3 hours per week, 25 - less than 3 hours per week, and 15 - never or hardly ever. 32 have been using a gamepad for at least 50% of their playtime, and 44 were playing (or used to play) mobile games. The most popular video game genres among participants were first-person shooters (FPS)/action (23 responses), flight simulation (22), racing (20), and strategy/management (20). Demographic, occupational, and personal characteristics distribution is presented in Tab. 4.1.

---

[3] One participant preferred not to say.

Table 4.1: Demographic, occupational, and personal characteristics distribution between the participants. The sample size was $N = 74$. FS - flight simulator; TS - touchscreen; VG - video games; hpw - hours per week; GP - gamepad. Where applicable: R - range, M - mean, SD - standard deviation, $\widetilde{m}$ - median.

Demographic characteristic	Distribution
Age (years)	R: 21 to 63, $M = 31.7$, $SD = 10.35$, $\widetilde{m} = 27$
Gender	Male: 55; Female: 18; Prefer not to say: 1
Handedness	Right: 64; Left: 7; Ambidextrous: 3
Fixed-wing total hours	R: 0 to 13300, $M = 1034$, $SD = 2477$, $\widetilde{m} = 4.5$
FS total hours	R: 0 to 1500, $M = 114$, $SD = 247$, $\widetilde{m} = 10$
TS attitude[4]	Answer (no. of responses): 1 (9); 2 (9); 3 (18); 4 (18); 5 (18)
VG frequency	yes, $>$ 3 hpw (13); used to, $>$ 3 hpw (22); yes, $<$ 3 hpw (13); used to, $<$ 3 hpw; no/hardly ever (15)
GP usage	a lot (12); sometimes (20); hardly ever (14); no (28)
Mobile games usage	yes (24); used to in the past[5] (20); no/hardly ever (30)
VG genres	FPS/action (23); flight sim (22); racing (20); strategy/management (20); indie/platformer (7); puzzle (7); RPG (5); sports (2)

4.3 Environment

The trials were conducted in the FSS, described in Chapter 3. This engineering flight simulator (EFS), developed as a part of this study, was made specifically to accommodate the research needs. The flight deck's touchscreen panels and authorship of the human-machine interface (HMI) code allowed the author to create a custom "trial operation widget", which included options to set the participant number, scenario, and inceptor. Furthermore, it had the ability to start and stop each trial attempt and monitor the simulation data. Each scenario was automated in a way that, after

---

[4] Question was "What is your view on touchscreen technology being introduced in aircraft cockpits? Use the scale from 1 (I do not like the idea) to 5 (I like the idea)". Two values are missing because the first two participants were not asked this question.

[5] In later parts of this thesis referred as "used to".

starting, it would automatically reset the aircraft's position, manage the aircraft's autopilot system, inject the disturbance signal, and record the data.

4.4 Inceptors

The design of the experiment required basic knowledge of how the aircraft is controlled. The elevator is situated in the aircraft's tail wings and is used for longitudinal control. By applying force to the stick or yoke by pushing or pulling, the pilot can control the pitch of the aircraft. Ailerons on the aircraft's wings control the aircraft's lateral movement – roll. The roll can be initiated by moving the inceptor sideways: left or right. Combined with pitch manipulation, this is the usual way of making turns in the business jet aircraft. The rudder is used to control the yaw. Since typically it is controlled by a pilot pressing the pedals, it was out of scope in this study. This was only the first attempt to introduce a new type of inceptor in the aircraft, and adding a third control surface to the analysis was found to be overwhelming and unnecessary; moreover, in good flight conditions, pilots only rarely use the rudder while in the air in this type of aircraft[6]. Overall, the control surfaces manipulate the lift and drag in different directions, giving control over the aircraft. Detailed physical analyses of that phenomenon were out of scope in this study and can be found in many engineering books such as *Aircraft Design: A Systems Engineering Approach* [305]. In fly-by-wire (FBW) aircraft, the inputs made by a pilot are usually translated into signals to the control system, which operates the elevator, ailerons and rudder. A similar principle is used in flight simulators: the physical input from the inceptor is translated into a digital signal and sent to the aircraft's dynamic model solver.

In this study, participants operated three different controllers in the FSS. The first was a traditional sidestick controller commonly used in business-type aircraft. The second was an Xbox gamepad, which is very familiar to video game console users. The third was a prototype touchscreen controller inspired by an on-screen joystick, found in some mobile games. They were implemented in the FSS and tested under various conditions to identify their strengths and limitations.

Aircraft sidesticks can be categorised as either "passive" or "active". The passive sidestick is designed to only provide resistance when physically pushed due to me-

---

[6] Conversation with a pilot who regularly flies Gulfstream business jets, personal communication, October 2019

chanical components like springs, dampers, and friction. Unlike an active sidestick, it cannot be moved without manually applying force. On the other hand, an active sidestick features technology that allows it to move independently of hand force, such as a motor [360]. The patent for the active sidestick technology is defined as a ”pilot flight control stick haptic feedback mechanism [which] provides variable force feedback to the pilot flight control stick” [150]. Both technologies have been utilised in flight simulators for years. Many military FBW aircraft have an active sidestick system, while civilian FBW aircraft, such as Airbus A380, only have passive sidesticks.

4.4.1 Sidestick

Sidestick is the default FSS controller, shown in Fig. 4.1. It can be classified as a passive displacement sidestick, meaning that the movement of the sidestick is directly proportional to the deflection of the control surfaces on the aircraft. It is a type of control input method commonly used in business jets and medium- to large-sized aircraft. For example, it can be found in the cockpits of the Airbus family (A320 to A380) and Dassault Falcon 7X. Studies have shown that there are only small differences in performance between active and passive sidesticks [129], and the number of cockpits equipped with passive sidesticks has increased significantly over the period of 2007 to 2017 [389]. As a result, the sidestick controller was an appropriate choice as a baseline for this study. Its passive nature kept the connection between the participant's input and stick deflection relatively simple [233]. The input control logic is presented

Figure 4.1: Future Systems Simulator's default inceptor - sidestick.

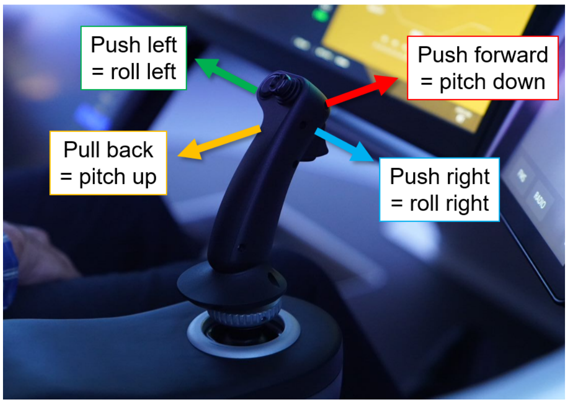


Figure 4.2: Sidestick's input logic.

in Fig. 4.2.

4.4.2 Gamepad

The Xbox gamepad is a gaming peripheral designed for use with the Xbox video game console, but it is also compatible with a PC through a regular USB cable. It consists of two analogue sticks, a d-pad, four action buttons, two shoulder triggers, two analogue triggers, and start and back buttons. The gamepad is ergonomically designed, with an asymmetrical shape that allows for comfortable grasp. The analogue sticks provide smooth and accurate control over a video game. It also offers vibration feedback; however, it was not implemented in this study, as other controllers did not have this feature. The gamepad was chosen to investigate how pilots with different gaming backgrounds would adapt and perform various tasks. The choice for this controller has been justified in Section 2.3.7. Participants were asked to hold the controller in both hands but only use their right thumb to move the right stick while resting both hands on the extendable tray in front of them, as presented in Fig. 4.3. The input control logic was the same as the sidestick's, as seen in Fig. 4.4.

Figure 4.3: Xbox gamepad controller.

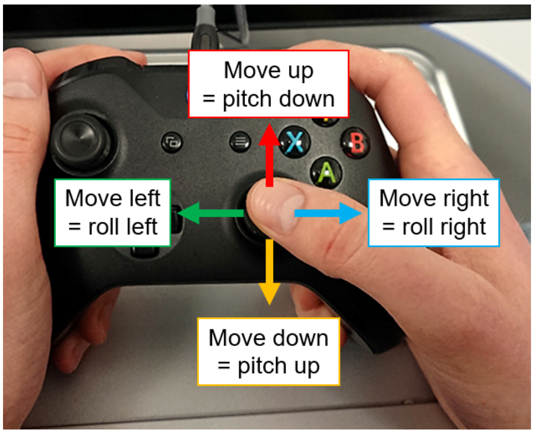


Figure 4.4: Gamepad's input logic.

4.4.3 Touchscreen

The controller developed specifically for this study was a modified mobile touchscreen displacement stick, also known as a "thumbstick". It was a touchscreen equivalent of a real joystick, allowing for a similar, "circular" movement using vertical and horizontal axes in a specified area. The design was based on interaction principles

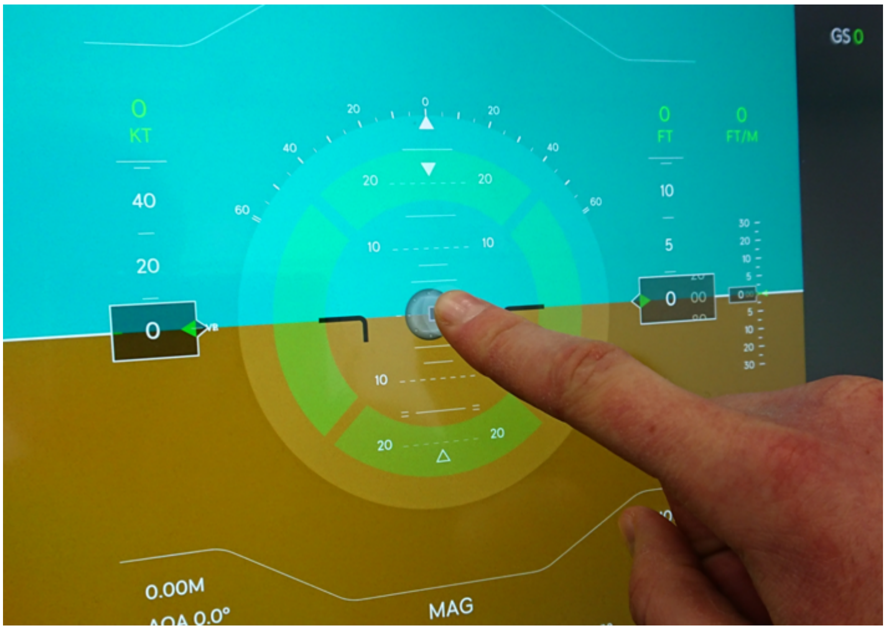


Figure 4.5: Novel touchscreen controller, adapted from mobile games' thumbstick.

similar to those of a real sidestick; thus, it was suitable to prototype it as an aircraft inceptor. The touchscreen controller was located in the middle of the attitude indicator on the primary flight display (PFD), on the FSS's right main display, as seen in Fig. 4.5. The display is fixed in the FSS's cockpit, so the user did not have to hold the monitor like a tablet to interact. Because of this, the user needed to ”tap and drag” to interact with the controller, which gave them a full range of movement within a specified area. This type of interaction can be described as ”casual prodder”, according to Telfer's *Touch Control Design* [345]. Some studies have shown that this type of controller, although lacking performance in video games like *Pac-Man* and *Super Mario Bros.*, is comfortable and intuitive [31]. The reason for the poor performance in those games might be because they are platformer-type of games which require a D-pad controller with an immediate digital response; this is in contrast with the touchscreen equivalent of analogue input, which is usually associated with flight simulators or racing games. There are many mobile games that use a touchscreen thumbstick as a control method. Some examples of these games include shooter-type games, such as *Fortnite Mobile*, *PUBG MOBILE*, or *Call of Duty: Mobile*, where the player uses a thumbstick to move the character around the game world and aim weapons; racing games, for example, *Asphalt 9: Legends* or *Real Racing 3* which use the touchscreen ”joystick” to steer the car, and also perform actions like braking and accelerating. Moreover, several flight simulator games for smartphones use a touchscreen thumbstick as the main control input. Some examples of these games include *X-Plane Flight Simulator*, *Infinite Flight Simulator*, or *Aerofly 2 Flight Simulator*[7].

Due to the software limitations at the time of conducting the trials, multi-touch was not implemented, hence only single-tapping and sliding gestures were used[9]. The participant interacted with the controller by touching the controller "stick" and, while holding it, moving it in any direction within the circle in the size of the attitude indicator. This technique is called "tap-n-drag", which was proven to be an intuitive interaction method for navigation in small-scale touchscreen devices [203]. The FSS's HMI application translates this input to inceptor signals and sends it directly to the aircraft model in a normalised range (-1 to 1) for vertical and horizontal channels. This particular research did not need to assure the accuracy of the touch; the proposed touchscreen controller was not in close contact with other touchscreen elements, hence initiating an interaction did not require high accuracy. If it was the case, there are several articles addressing the optimal size of touchscreen elements. Pahri et al. specified the optimal size for thumb-operated touchscreen buttons to be 9.6mm-wide for discrete (single-target - buttons) tasks / 7.7mm-wide for serial (sequence of taps – text entry) tasks [278]. Tao et al. presented results on different button layouts, with the best results obtained for 17.5mm and larger square buttons [341]. Xiong et al. also recommended that a square button should be at least 10-12mm wide [394].

Contrary to the sidestick and gamepad, the touchscreen controller had a reversed vertical axis, as seen in Fig. 4.6. In this case, when the pilot moved the finger up, the aircraft pitched up (while with the sidestick and gamepad, the aircraft pitched up when the pilot pulled the stick back). The vertical channel was reversed because initial trials conducted within the author's research group showed that having that channel the other way would confuse the participants even more. In the

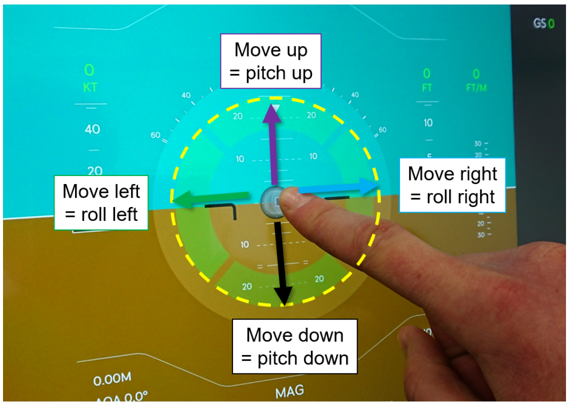


Figure 4.6: Touchscreen controller's input logic.

---

[7] All listed mobile games (apart from *Fortnite Mobile*) can be found on Google Play Store for Android OS[8] (play.google.com, accessed on 2022-12-22). *Fortnite Mobile* can be downloaded from Epic Games website (epicgames.com, accessed on 2022-12-22)

[9] The touchscreen panels of FSS are multi-touch-ready, so this feature will be added in a future upgrade.

current state, the aircraft goes in the di-
rection pointed by the user. Being positioned in the centre of the PFD, this philosophy was the most intuitive option, even though it caused some confusion among participants with flight experience. Kivila [192] noted that the "stimulus-response compatibility", a term first used by Fitts & Seegar [127, 128], applies to professional users when introducing control methods radically different than those already known and may create a bias when a change is introduced. During trials for this study, it was observed that the inverted Y-axis mapping usually caused more errors among participants with some aerial experience as compared to those who had never flown any aircraft. The inverted Y-Axis appeared to be more intuitive and easy to learn for naïve participants, which is apparent in the results Section 5.7.3. Moreover, studies by Kwon et al. and Zhang et al. showed that, in some cases, inverting the direction of "touch-n-drag" and "slide" techniques did not induce any negative effect of the direction in touchscreen control [203, 398].

Further analysis of this behaviour can be described by a control display gain (CDG), which is a relationship ratio between the input device (here: participant's finger) and the display pointer (here: a touchpoint on the touchscreen controller) [136, 311]. CDG is a unitless coefficient, often described by levels, which is the relation of input to output (display). For example, level 1 signifies that the user's input movement translates in 1:1 relation to the pointer on display, while level 3 means that the movement speed of a pointer is tripled in relation to the input signal. CDG was studied with various input devices: computer mouse [61], gestures [289], distal/ray-cast pointing using an air mouse [70], and touchscreen [203]. Casiez et al. noted that the control device and display pointer are often decoupled. They defined two spaces: (1) "the display space" - a representation of the pointing action, and (2) "the motor space" - a manipulation of the control device [61]. In this research, the aircraft's flight dynamics model and its response to the pilot's input was the motor space, which was out of scope. The findings of previous CDG studies were that, for mouse control, higher gain increases performance (with gain at level 4 being participants' preference for a regular-sized screen) [61]; for gestures, lower gain (close to level 1) showed the best results [289]; with ray-cast pointing, it was proven that CDG level 1 or less gives the best Usability results [70]; gain on levels 3 and 5 for touchscreen navigation on a mobile phone screen were the easiest for participants; however, levels 1 and 3 were best for fine control [203] and gain equal and close to level 1 showed best results in "drag-and-drop", and rotation tasks in vibration environments [342]. Overall, higher gain tends to be better for "detached" control devices such as a mouse or gestures and for touchscreen for screen navigation; however, tasks that require precision tend to be completed faster at lower gains. The touchscreen controller was mapped in a way that it sent the input signal to the model in the same way that the sidestick did - in a linear range from -1 to 1 in both the horizontal and vertical axis. The ratio between the input device and the display pointer (CDG) was at level 1 because the touch area was relatively small, and it did not require a user to move the pointer between the screens. This ensured that the control pointer was always positioned exactly where the user's finger was. The inverted Y-axis gave the user the expected action from the input (pointing up meant pitching up; pointing down meant pitching down), which complies with the "Input Equals Output" theme in Organic User Interfaces design [368].

4.5 Tasks

The study included four tasks designed to evaluate the participant's performance in different scenarios. The goal of each of them was to maintain control of the aircraft and to accurately track a desired flight path or trajectory. All four tasks were repeated for each of the three inceptors presented in the previous section.

The first two tasks were manual-control disturbance rejection (DR) tasks, which are also known as "compensatory disturbance tracking tasks" [365]. They involved controlling the aircraft in the presence of external disturbances, such as wind gusts or turbulence, but only on one control channel at a time. The first task, disturbance rejection in the vertical channel (DRV), tested the participant's ability to maintain control of the aircraft's pitch and altitude; the second task, disturbance rejection in the horizontal channel (DRH), checked the participant's capacity to maintain control of the aircraft's roll.

The other two tasks were manual-control landing tasks, which tested the pilot's ability to land the aircraft safely and accurately under different conditions. The first was a landing with no disturbance (LN), which involved landing the aircraft in clear conditions without any disturbances. The second task, landing with disturbance (LD), was to land the aircraft in the presence of external turbulence in both channels, vertical and horizontal. The pilot had to compensate for this disturbance to maintain a straight flight path.

All the scenarios were conducted under clear weather conditions, with no wind or other air or airport traffic. However, some of the scenarios involved the introduction of simulated turbulence as a disturbance. The disturbance signal for each task was generated using a pre-defined sum-of-sines forcing function, using Lone's MATLAB function [233]. This signal simulated a "continuous turbulence" of an aircraft, as opposed to screen vibration or a full simulation motion. It was introduced to excite the participant's controller movement and measure their performance when compared to the goal or baseline. The disturbance signal was different for each task but kept consistent between participants to allow exact comparison in later analyses.

The scenarios chosen for this study were designed to challenge and engage the participants and evaluate their performance using different controllers. The disturbance tasks were explicitly chosen to allow the participants to familiarise themselves with control in only one channel at a time and then to test their ability to use both channels simultaneously in a landing scenario, with and without added disturbance. This study focused specifically on the experience of pilots while interacting with pitch/roll inceptor, so all other cockpit elements, such as engine throttle, landing gear, flaps, spoilers, brakes, and autopilot control, were automated in order to reduce the workload of the participants. Additionally, participants were instructed to minimise the use of the rudder pedals while in flight, as it is a usual practice in business jet aircraft[10].

It was determined that the order of the scenarios should not be random, especially in a case where non-pilots were involved. The chosen order of the scenarios imposed a small learning curve to allow for a better chance of success. For example, if the landing scenario with disturbance was given first, without any prior practice with the DR tasks, the likelihood of failure would be much higher.

4.5.1 Disturbance rejection tasks (DRV and DRH)

The DR tasks were divided into two sub-scenarios: one in the vertical channel (also known as the pitch/lateral axis) and one in the horizontal channel (also known as the roll/longitudinal axis).

---

[10] Conversation with a pilot who regularly flies Gulfstream business jets, personal communication, October 2019

DRV: Disturbance rejection task (vertical channel)

The disturbance rejection task in the vertical channel (DRV; about the lateral axis) was the first scenario in this study. During the trial period, the task was revised twice. The first revision occurred after the first two participants, as it became apparent that most non-pilots might struggle with the task due to the low initial altitude. The disturbance in pitch control caused the aircraft to descend rapidly, and the participants did not have a way to ascend back to the initial altitude. After the task was revised and the aim changed to keep the flight vector on the horizon line, participants 3-10 were able to maintain a relatively reasonable altitude. However, it was observed that those participants, despite being experienced line pilots, still ended up at very low altitudes. A further task revision was deemed necessary in order to make the task suitable for less experienced subjects. A detailed description of the task revision motivation and the process can be found in the following subsection, *DRV task revision*.

For the first two participants, the task involved keeping the pitch attitude indicator at the centre of the PFD, as seen in Fig. 4.7. For participants 3-10, the task involved keeping the flight vector on the horizon line of the PFD (maintaining a level flight; Fig. 4.8). For all remaining participants, the task involved acquiring and keeping the target altitude of 1500 feet, shown in Fig. 4.9. In all cases, the aircraft was only moving about its lateral axis.

Each manoeuvre began with a 10-second countdown, during which the autopilot control was enabled. After the countdown, the participants gained control of the aircraft. After a few seconds, the disturbance was introduced to the inceptor's pitch channel, and the participants were required to compensate to meet the specific task requirements described above. After several seconds, the disturbance ended, and the participants had a few additional seconds to level and stabilise the aircraft. The end of the scenario was signalled with a "ping" sound, accompanied by a message on top of the PFD.

For this task, the roll control channel was disabled, so the participants could only move the aircraft up and down. The zero-degree roll was kept by autopilot in the horizontal channel (longitudinal mode).

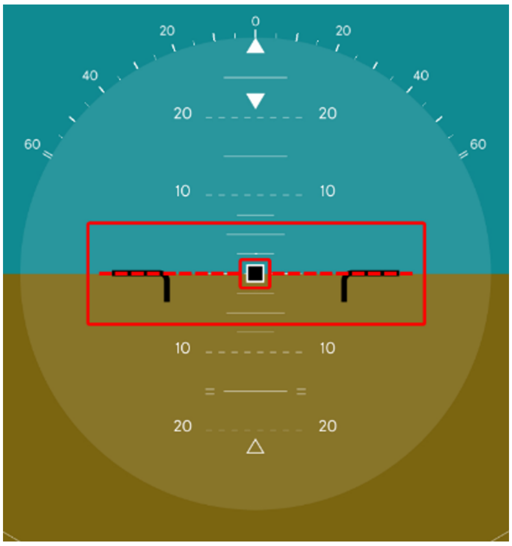


Figure 4.7: A visualisation from the pre-flight briefing of the first iteration of DRV scenario task, performed by participants 1 and 2 in the pilot study. The aim was to keep the aircraft's pitch (black dot with a white outline on the centre of the PFD) on the artificial horizon.

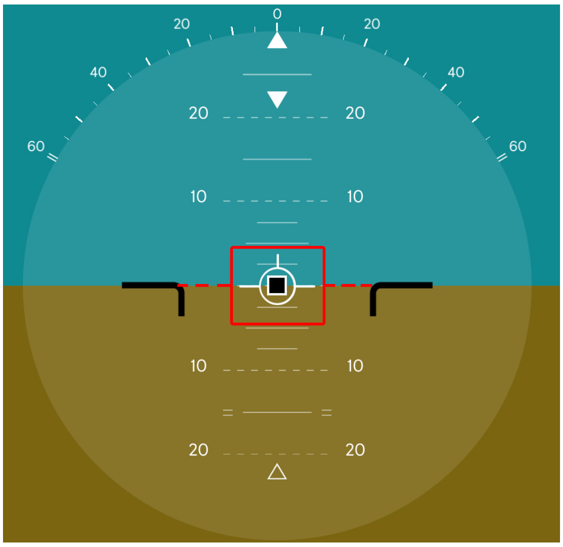


Figure 4.8: A visualisation from the pre-flight briefing of the second iteration of DRV scenario task, performed by participants 3-10 in the pilot study. The aim was to keep the aircraft's flight path vector (a white plane symbol on the centre of the PFD) on the artificial horizon.

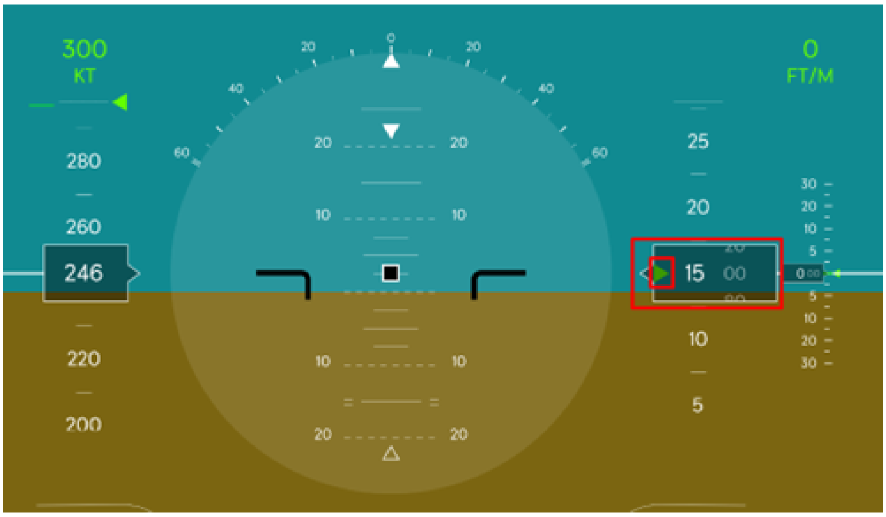


Figure 4.9: A visualisation from the pre-flight briefing of the third iteration of DRV scenario task, performed by participants 11+. The aim was to capture and hold the target altitude of 1500 feet, marked by the green triangle on the altitude indicator.

DRV task revision

The initial design of the DRV task was found to have some flaws during the pilot trials. Specifically, the implemented disturbance signal caused the aircraft to pitch down suddenly, plummeting towards the ground. Because of the task definition, they were unable to get back up to the initial altitude level. As a result, pilots were unable to maintain a comfortable altitude and experienced stress, as indicated by their post-trial feedback[11] and questionnaire responses. In some cases, pilots had to pitch up in order to avoid crashing to the ground, which was not in line with the task aim of keeping the pitch indicator on the horizon line. Changing the initial altitude of the aircraft was not a viable option because it would require the recompilation of the aircraft's flight dynamic model. Due to the equipment limitations, a single recompilation would take around 5 minutes. Between DR and landing scenarios and for every inceptor, this would have prolonged the trial schedule for each participant by 30 minutes, thus overflowing the experiment's time constraints.

In order to address this issue, the DRV task was revised to require pilots to maintain the flight vector on the horizon line instead of the pitch attitude indicator. However, it was found that some participants still experienced difficulties maintaining altitude due to the nature of the disturbance. As a result, the DRV task was revised again to focus on maintaining a specified target altitude, with desired and acceptable error margins communicated to the participant before each scenario. This change aimed to help participants focus more on inceptor handling and reacting to disturbance changes, as well as to provide a more meaningful measure of performance and potentially reduce stress caused by close proximity to the ground. A visualisation from the pre-flight briefing, presenting the task goal, is shown in Fig. 4.9.

DRH: Disturbance rejection task (horizontal channel)

The second scenario was a disturbance rejection task in the horizontal channel (DRH). The aim was to keep the roll indicator in its centre position to maintain a zero-degree roll of the aircraft, as shown in Fig. 4.10. This scenario was only limited to the longitudinal axis, and it did not need revisions, as the altitude was kept constant by the autopilot in the vertical channel (lateral mode).

---

[11] One of the participants wrote *"Proximity to terrain increases workload/discomfort during tasks"*, relating to the DRV task.

The manoeuvre was similar to the DRV task: it began with a 10-second countdown, during which the autopilot control was enabled. After the countdown, the participants gained lateral control of the aircraft. After a few seconds, the disturbance was introduced to the inceptor's roll channel, and the participants were required to compensate by keeping the roll indicator on the centre position of the PFD. After several seconds, the disturbance ended, and the participants had a few additional seconds to level or stabilise the aircraft.

For this task, the pitch control channel was disabled, so the participants could only move (roll) the aircraft left and right.

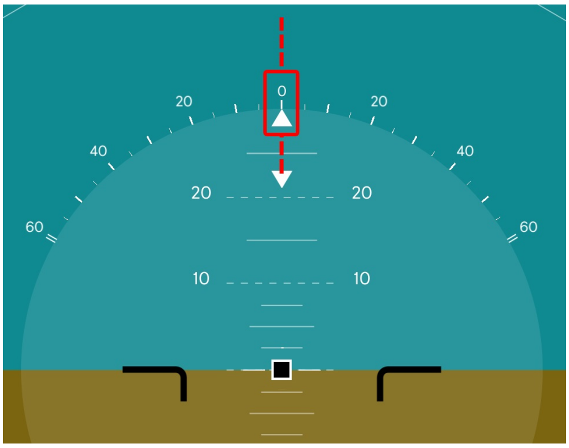


Figure 4.10: A visualisation from the pre-flight briefing of the DRH scenario task. The aim was to keep the aircraft in a zero-degree-roll flight, indicated by an upwards-pointing white triangle of the PFD's roll indicator.

4.5.2 Landing tasks (LN and LD)

The landing scenarios in this study were designed to evaluate the participant's ability to control the aircraft with a specific controller and successfully land it under different conditions. There were two types of landing tasks: one in clear conditions (”Landing with No disturbance”, LN) and one with additional simulated turbulence (”Landing with Disturbance”, LD).

Initially, the landing tasks were offset-landing-type scenarios, where the aircraft was not positioned in a straight line to the runway, and the pilot had to manoeuvre the aircraft to align with the runway first. However, after pilot trials, it was determined that these tasks would be too challenging for less experienced and naïve pilots, so for participants 11 and up, the landing tasks were revised to be straight landings without the need to perform a turn manoeuvre. The detailed justification can be found in the next section.

For the first 10 participants, the starting position of the aircraft was 5 miles away from runway 15R at Incheon International Airport in Seoul, South Korea[12], with an initial speed of 150 knots (0.25 Mach), an altitude of 1400 ft, and a heading of 90 degrees, shown in Fig. 4.11 and 4.12. The desired glide slope was 3 degrees. To align with the runway, the pilot had to execute a 54.5-degree turn manoeuvre approximately halfway through the approach, using the visual cues and flight deck elements such as a map on the navigation display (ND), the flight path vector indicator and the instrument landing system (ILS)[13] on the PFD.

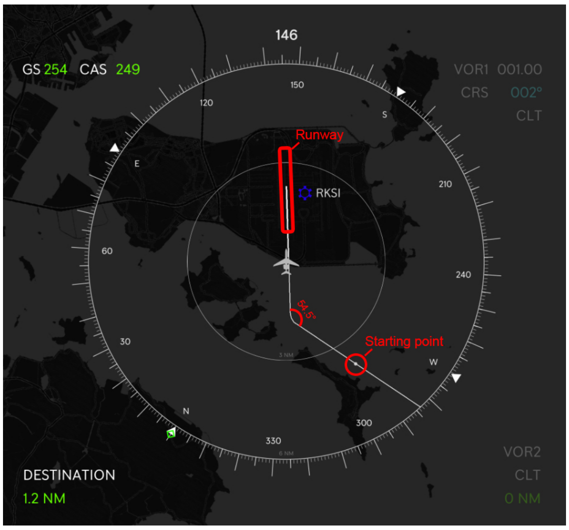


Figure 4.11: A visualisation from the pre-flight briefing of the first iteration of landing tasks, performed by participants 1-10 in the pilot study (offset landing). During the descent, the pilot had to make a turn manoeuvre approximately halfway through the approach to align with the runway. The figure shows the Navigation Display of the FSS, with task information marked in red.

Figure 4.12: A satellite image (taken from maps.google.com) section from the pre-flight briefing of the first iteration of landing tasks, showing the approximate flight path approach.

For the remaining participants, the starting position was 5 miles away from the runway in a straight line, with an initial altitude of 1500 ft and a heading of 144.5 degrees, presented in Fig. 4.13 and 4.14. The rest of the task conditions remained the same.

---

[12] This particular airport was chosen because the aircraft navigation systems were optimised for this location for earlier trials carried out in the FSS.

[13] In the offset-landing task, the horizontal ILS was only usable after the turn. This was due to the model limitations, but it was communicated to the participants. Straight landing task did not have this issue.

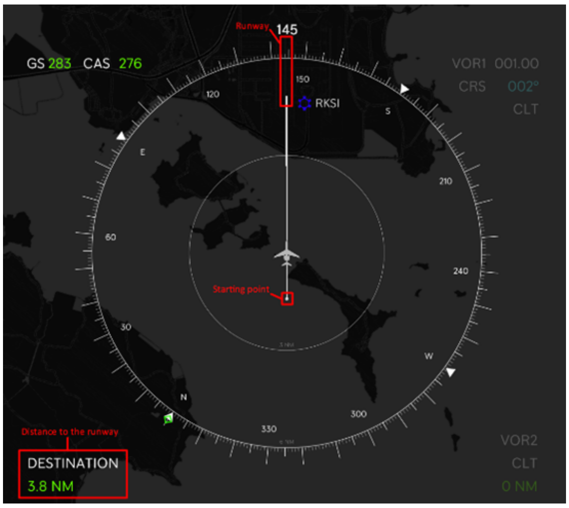


Figure 4.13: A visualisation from the pre-flight briefing of the second iteration of landing tasks, performed by participants 11+ in the main study (straight landing). During the descent, the pilot had to keep the aircraft aligned with the runway while maintaining the correct flight path. The figure shows the Navigation Display of the FSS, with task information marked in red.

Figure 4.14: A satellite image (taken from maps.google.com) section from the pre-flight briefing of the second iteration of landing tasks, showing the approximate flight path approach.

In both LN and LD scenarios, there was a 10-second countdown at the beginning of the task, during which autopilot control was enabled. After the countdown, the pilot gained control of the aircraft. 5 seconds later, in the LD scenario, a disturbance was introduced to the inceptor's lateral and longitudinal channels, which the participant had to compensate for in order to maintain the desired flight path. The disturbance lasted for 90 seconds, which was the time it took for the aircraft to almost reach the runway. If the participant felt that a go-around was necessary at any point during the approach, they were asked to announce it, but to still continue the approach and try to land.

Flaps, landing gear, and autothrottle (at a target speed of 120 knots) were pre-set and maintained until close to the runway. Right before touchdown, the autothrottle automatically disengaged, the spoilers were extended, and the brakes were engaged. The threshold for this automation was 100 ft for the first 11 pilots and was revised to 50 ft for the remaining participants. The automation of these elements was implemented to reduce the training time for non-pilots and to decrease workload, allowing them to focus solely on inceptor control. The end of the scenario was signalled with a ”ping” sound and a message on the PFD.

Landing tasks revision

The LN and LD scenarios were found to be very challenging for pilots, particularly when using unfamiliar inceptors. Most pilots were able to land the aircraft successfully; however, in many cases, the downward G-force would result in a simulated crash in non-realistic conditions. Non-pilots who were not experienced with flying at all would likely have struggled to complete these tasks. Previous research conducted by the author in 2017/18 (unpublished) revealed that some non-pilots had difficulty landing the aircraft even when flying in a straight line towards the runway and without any disturbance. Furthermore, feedback from the pilot study on usability, workload, and situation awareness (SA) questionnaires indicated that the landing scenarios, particularly LD, were extremely physically and mentally demanding. While this level of challenge may be suitable for other types of experiments, the goal of this study was to focus on inceptor handling without overloading the pilots. To ensure that these tasks were manageable for all participants, it was decided to revise them.

Therefore, the LN and LD tasks were modified to remove the requirement for pilots to turn before aligning with the runway and perform a straight landing instead. This meant that the starting position of the aircraft was in a straight line to the runway, allowing pilots to focus more on maintaining a steady glide slope and alignment while still being challenged by the presence of disturbance. This revision also enabled the horizontal channel of the ILS to work correctly from the outset and eliminated the risk of ”overshooting”, which happens when a pilot turns the aircraft too late. The difference can be seen between Fig. 4.11 & 4.12 and 4.13 & 4.14

In addition to changing the initial conditions of the landing tasks, the auto-braking threshold was also revised after the 11th participant suggested that it was activated too quickly, reducing the pilot’s ability to make a precise landing. Therefore, the threshold was changed from 100 to 50 feet. Although it was still unrealistic for experienced pilots, it helped participants with no experience land the aircraft.

4.6 Trial procedure

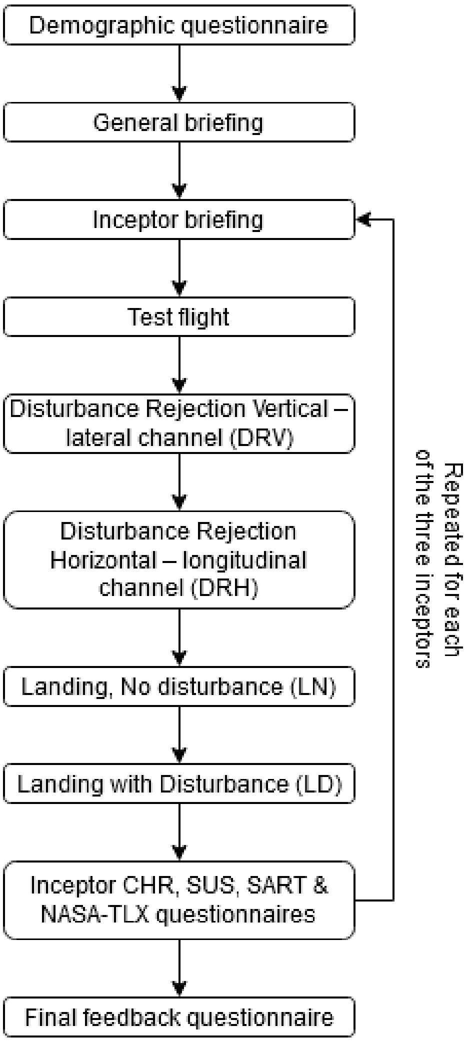


Figure 4.15: Diagram showing the trial procedure.

The trial procedure (presented in Fig. 4.15) began with the participant signing the consent form. The demographic questionnaire was completed prior to the arrival. Following that, they were presented with an 8-minute briefing. After the briefing video, the participant was seated in the simulator, and the video recording was started. In some cases, before starting the experiment, a heart rate variability (HRV) monitor was attached to the ear, and an eye tracker was given to the participant, followed by the devices' short calibration (depending on the personnel availability). In the meantime, the simulator's HMI was set up: the autopilot was engaged, heading and altitude targets were set to respectively 144 degrees and 1500 feet, the synoptics page on the FSS's central lower display was hidden to decrease distraction, and the inceptor and scenario were set. The participant was allowed to adjust the seat and the sidestick pedestal, and then they were shown printed information about the inceptor they would be using. The order of the inceptors was randomised. If it was a gamepad, they were asked to keep both hands on the extendable tray; with a touchscreen, they were instructed how to interact with the controller, as it was limited to only one touch at a time. Moreover, the controller did not recognise touch if it was read by the Windows interface as a right click – the pilot had to slightly "slide" the finger upon touching the screen instead of holding it in place. Additionally, they were reminded that the Y-axis of the touchscreen controller was inverted. After the preparations, the pilot was given one minute of a free test flight without any tasks. This was meant for them to familiarise themselves with the aircraft "feel" and the steering method. After that, the scenarios were initialised, one after another: DRV, DRH, LN and LD. Before the first two, the participant was informed about the respective targets and desirable and acceptable margins. For DRH, the participant was reminded that the target is to hold the roll position, not the heading. For LN, the participant was reminded to switch from using the ILS to PAPI lights[14] at 1.5 miles to the runway, and they were informed that the automatic breaking would engage at 50 feet above the ground, so the throttle lever and some elements of the central upper display would move, such as the spoilers lever, to decrease the speed. They were also reminded to ignore the green triangle bug on the altitude indicator (target altitude for autopilot). After completing all tasks with one inceptor, they were given a questionnaire with CHR, SUS, SART and NASA-TLX scales. The rating was done for each scenario. After that, two other inceptors were tested in the same manner as the first one, followed by a questionnaire with rating scales. When this was done, the participant was given a final questionnaire with an open feedback opportunity (qualitative measure). In the meantime, the HRV monitor was removed, and the video recording stopped. The whole process took approximately 1 to 1.5 hours.

4.7 Collected data

During each session, the following sets of data were collected:

1. Demographic information, including the participants' experience with piloting aircraft and their attitude towards the introduction of touchscreen technology in flight decks.
2. The participant's control input data, both raw and with added disturbance.
3. Flight simulation data, including the aircraft's aerodynamics.
4. Subjective ratings on the inceptor's usability and handling qualities (HQ), and pilot's perceived workload and SA.
5. Video recordings of the participant (top-view) for additional single-case manual analyses, such as additional verbal feedback or a go-around call confirmation.

Demographic and experience information was collected in order to classify the participants and investigate various grouping factors as independent variables. This was

---

[14] Precision approach path indicator (PAPI) lights are visual aids located at the side of the runway at an airport. They provide an indication of the aircraft's position relative to the desired approach path and assist pilots in maintaining the correct flight path during a landing approach. PAPI lights consist of four lights: red colour indicates that the aircraft is below the proper glide slope, while white colour indicates that the aircraft is above the flight path. Ideally, they should show two white and two red lights.

done to assess if any of them had a significant impact on the results from the trials or interaction with any of the inceptors. The simulator was programmed to record the flight and input data for each inceptor and scenario combination. Recorded variables included the aircraft's position, forces, rates of movement and rotation, navigation systems data, autopilot and autothrottle signals, pilot's inceptor inputs and the implemented disturbance signal. Depending on the availability of the devices and personnel, additional data collected for some of the participants included heart rate variability measure, recorded with HeartMath's *Inner Balance Coherence Sensor* and *Inner Balance App* [222] and eye-tracking data, collected using Pupil Labs' "Pupil Core" eye tracker [286].

4.7.1 Handling qualities assessment

Flying qualities (FQ), flying performance and flying systems are three principal regimes in the flight testing field of science. FQ involves broad research and assessment of the control characteristics of any given aircraft. They play a major role in assuring the safety of a flight and the easiness of controlling the aircraft during steady flight and regular manoeuvres [139]. Handling qualities (HQ) are a more precise term than FQ – they relate to an aircraft's stability and characterise the level of difficulty and precision with which a pilot performs a task in flight [85]. HQ has seen a significant amount of research for many years, starting from G. H. Bryan in England in 1904, who defined the theory of stability of aeroplanes. There are standards that define "good HQ" for any given aircraft, be it a fixed-wing [15] or a rotor-wing aircraft [38], and its specific flying task. The HMI is one of the aspects of the HQ. HQ have been in development since the first aircraft has come into existence, especially since the 1970s [255]. It is a never-ending process of proposing, testing and optimising at every stage, from control devices through the technology (e.g. FBW), up to the control surfaces (e.g. ailerons, elevator, rudder) and pilot's neuromuscular responses [27, 72, 76, 318]. During the development and research of existing and new aircraft technologies, HQ should not only be assessed by mathematical analyses but also through the pilot's feedback and commentary, according to Cooper and Harper [151]. There are several ways for the test pilot to assess the experience. This research utilised four methods: Cooper-Harper Rating Scale (CHR) [85], System usability scale (SUS) [54, 245], Situational awareness rating technique (SART) [344], and NASA-TLX (Task Load Index) [155], presented below.

4.7.2 CHR: Cooper-Harper Rating Scale

The Cooper-Harper Rating Scale (CHR) is a tool used to assess the workload of pilots during flight [85]. It consists of a 10-point scale that ranges from 1 (very low workload) to 10 (very high workload). The scale is based on the pilot's subjective rating of the workload associated with a given task. The pilot rates their workload by answering questions on the diagram (shown in Fig. 4.16), starting from the bottom left corner. The questions lead the rater to a number from the scale that best reflects their perceived experience. The CHR can be used in combination with other workload assessment tools, such as the NASA Task Load Index (NASA-TLX), to provide a complete understanding of the workload experienced by pilots during different phases of flight. The scores 1-3 mean low workload, 4-6 show high workload, and 7-10 suggest very high workload.

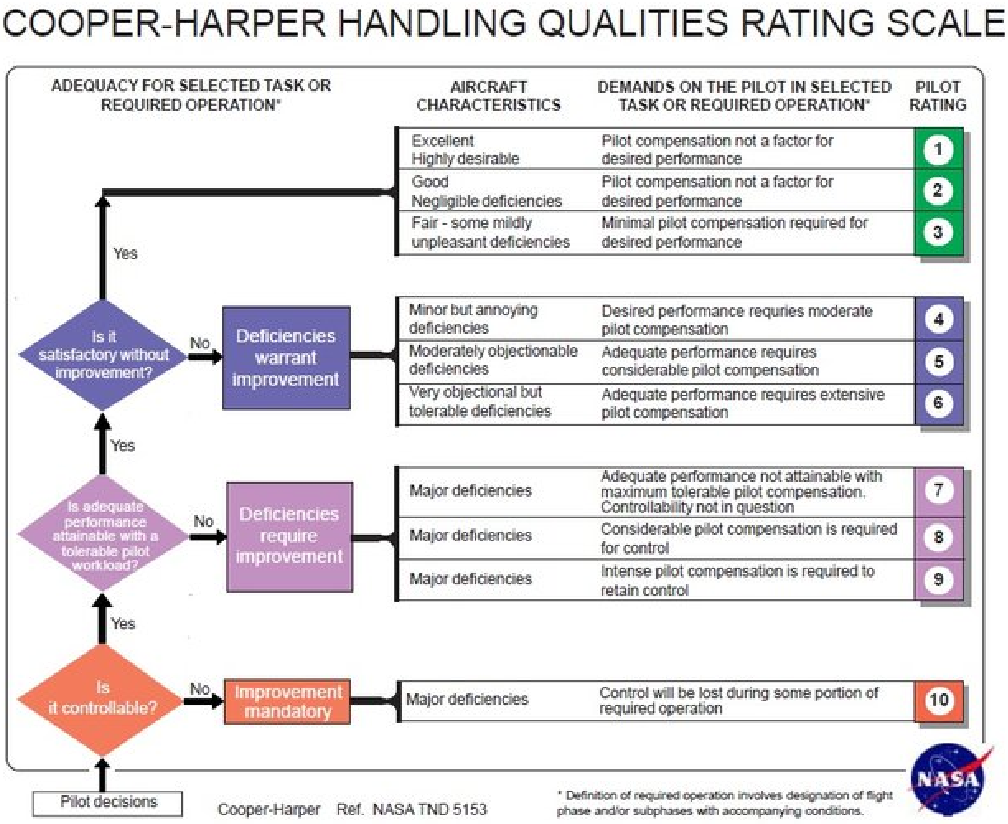


Figure 4.16: Cooper-Harper Rating Scale diagram. Reproduced from Cooper & Harper [85].

4.7.3 SUS: System usability scale

The system usability scale (SUS) is a standardised questionnaire that is used to measure the usability of a product or system [54]. It consists of 10 statements that participants are asked to rate on a 5-point Likert scale, ranging from ”strongly disagree” (1) to ”strongly agree” (5). The scores for each statement are then added and transformed into a scale from 0 to 100, with higher scores indicating higher usability. The statements are:

1. ”I think I would like to use this control system frequently”.
2. ”I found this control system unnecessarily complex”.
3. ”I thought this control system was easy to use”.
4. ”I think that I would need the support of a technical person to be able to use this control system”.
5. ”I found the various functions in this control system were well integrated”.
6. ”I thought there was too much inconsistency in this control system”.
7. ”I would imagine that most people would learn to use this control system very quickly”.
8. ”I found this control system very awkward to use”.
9. ”I felt very confident using this control system”.
10. ”I needed to learn a lot of things before I could get going with this control system” [54].

The overall SUS score (SUS-Total) can be further decomposed into two subdimensions: usability (SUS-U) and learnability (SUS-L). Usability is defined as the ease with which a user can learn and use the system. It makes 80% of the total scale score, while learnability is the rest 20% and is defined as the ease with which a user can learn to use the system. The SUS-Total, SUS-U and SUS-L scores are calculated as follows:

$$\text{SUS-Total} = \text{SUS-U} + \text{SUS-L}, \tag{4.1}$$

$$\text{SUS-U} = (Q_1-1+5-Q_2+Q_3-1+Q_5-1+5-Q_6+Q_7-1+5-Q_8+Q_9-1)*2.5, \tag{4.2}$$

and

$$\text{SUS-L} = (5 - Q_4 + 5 - Q_{10}) * 2.5, \tag{4.3}$$

where Q_n is an answer to the n-th question from a SUS questionnaire.

The SUS scoring can be interpreted in several ways. Bangor et al. developed two descriptions: "acceptability ranges" [35] and "adjective ratings" [36]. However, those scores only refer to the SUS-Total. The author could not find the interpretation of the SUS-U and SUS-L subdimensions, so the scaled versions of acceptability ranges and adjective ratings were used. The re-scaled ranges can be seen in Tab. 4.2.

Table 4.2: SUS interpretation for adjective ratings (upper part) and acceptability ranges (lower part), adapted for SUS subdimensions and based on Bangor et al. [35, 36] and Sauro [315].

	SUS-Total		SUS-U		SUS-L	
Scale	$\geq$	$>$	$\geq$	$>$	$\geq$	$>$
Worst	0.0	25.0	0.0	20.0	0.0	5.0
Poor	25.0	38.0	20.0	30.4	5.0	7.6
OK / Fair	38.0	52.0	30.4	41.6	7.6	10.4
Good	52.0	73.0	41.6	58.4	10.4	14.6
Excellent	73.0	85.0	58.4	68.0	14.6	17.0
Best	85.0	$\leq$100.0	68.0	$\leq$80.0	17.0	$\leq$20.0
Not acceptable	0.0	50.0	0.0	40.0	0.0	10.0
Marginal (low)	50.0	62.0	40.0	49.6	10.0	12.4
Marginal (high)	62.0	70.0	49.6	56.0	12.4	14.0
Acceptable	70.0	$\leq$100.0	56.0	$\leq$80.0	14.0	$\leq$20.0

4.7.4 SART: Situational awareness rating technique

The situational awareness rating technique (SART) is a subjective measure of situation awareness (SA) [344]. SA refers to the ability of an individual to comprehend the current situation and its potential future developments and to use this understanding to anticipate potential problems and take appropriate actions [156].

SART consists of three subdimensions: demand, supply, and understanding. The demand subdimension (SART-D) measures the degree to which the task or situation requires mental effort, concentration, or information processing. The supply subdimension (SART-S) measures the degree to which the individual has the necessary information, resources, and tools to perform the task or cope with the situation. The understanding subdimension (SART-U) measures the degree to which the individual understands the current situation and its potential future developments.

To use SART, the participant is asked to rate their situation experience on a 7-point scale for the following of questions, with 1 representing "low" and 7 representing "high":

1. "Instability of Situation: How changeable was the situation? Was the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)?"
2. "Complexity of Situation: How complicated was the situation? Was it complex with many interrelated components (High) or was it simple and straightforward (Low)?"
3. "Variability of Situation: How many variables were changing within the situation? Were there a large number of factors varying (High) or were there very few variables changing (Low)?"
4. "Arousal: How aroused were you in the situation? Were you alert and ready for activity (High) or did you have a low degree of alertness (Low)?"
5. "Concentration of Attention: How much were you concentrating on the situation? Were you concentrating on many aspects of the situation (High) focused on only one (Low)?"
6. "Division of Attention: How much was your attention divided in the situation? Were you concentrating on many aspects of the situation (High) or focused on only one (Low)?"
7. "Spare Mental Capacity: How much mental capacity did you have to spare in the situation? Did you have sufficient to attend to many variables (High) or nothing to spare at all (Low)?"
8. "Information Quantity: How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?"
9. "Information Quality: How good is the information you have gained about the situation? Was the knowledge communicated very useful (High) or was it insufficient (Low)?"
10. "Familiarity with Situation: How familiar were you with the situation? Did you have a great deal of relevant experience (High) or was it a new situation (Low)?" [344]

The total SART score (SART-Total) is calculated by subtracting the difference between SART-D and SART-S from SART-U. The equations 4.4-4.7 present the calculations for the SART-Total and the subdimensions scores, and they are as follows:

$$\text{SART-Total} = \text{SART-U} - (\text{SART-D} - \text{SART-S}), \tag{4.4}$$

$$\text{SART-D} = Q_1 + Q_2 + Q_3, \tag{4.5}$$

$$\text{SART-S} = Q_4 + Q_5 + Q_6 + Q_7, \tag{4.6}$$

and

$$\text{SART-U} = Q_8 + Q_9 + Q_{10}, \tag{4.7}$$

where Q_n is an answer to the n-th question from a SART questionnaire.

The scale is not normalised and ranges from -21 to 46 for the total score, 3 to 21 for demand and understanding subdimensions, and 4 to 28 for the supply subdimension.

There is no specific reference that provides guidelines on how to interpret SART scores because it is a subjective construct that refers to an individual's perception of the elements in their environment, the actions they can take, and the potential consequences of those actions. Therefore, it is crucial to consider the context in which SART is being used and the specific goals of the assessment when interpreting the scores [299].

One approach to interpreting SART scores is to consider the scores in relation to the mean or median scores among participants. For example, scores that are above the average may be considered to indicate higher levels of SA. Respectively, scores below the average will indicate lower SA. Another approach can be defined as follows: scores above 70% indicate high SA, 50% to 70% suggest moderate SA, and below 50% can be interpreted as a low SA; however, these categories are not strict, and the interpretation of SART scores may vary depending on the specific context and task. By combining those two approaches, three categories were defined by calculating thresholds based on mean scores for each dimension, presented in Tab. 4.3.

It was also important to consider the individual subdimensions of SART, as they provide additional insight into an individual's SA. For example, a high score on the demand subdimension may indicate that the individual was experiencing high levels of workload or stress, which could potentially impact their SA. Similarly, a low score on the supply subdimension may indicate that the individual has not received enough information to effectively monitor and understand the situation.

Table 4.3: Categorisation of SART scores. The mean percentage value M was rounded to a multiple of 10. Based on that, the low ($M - 10\%$) and high ($M + 10\%$) thresholds were calculated (and rounded to whole numbers). The "moderate" range is an open interval (does not include endpoints).

	SART-D	SART-S	SART-U	SART-Total
Mean	11.56	18.40	14.30	21.14
Mean (%)	≈ 50	≈ 60	≈ 60	≈ 60
Low	≤ 10	≤ 16	≤ 12	≤ 12
Moderate	10-14	16-21	12-16	12-26
High	≥ 14	≥ 21	≥ 16	≥ 26

4.7.5 NASA-TLX: Task Load Index

NASA's *Task Load Index* (NASA-TLX) is one of the most common ways to measure workload index. Developed in the 1980s, it is still widely used among researchers. It consists of six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Workload score can be interpreted as follows: low (0-9), medium (10-29), somewhat high (30-49), high (50-79) and very high (80-100) [155]. Each dimension is measured by one question on a 7-point scale (from 1, "very low", to 7, "very high") as follows:

1. "Mental Demand: How mentally demanding was the task? How much mental and perceptual activity was required? Was the task easy and simple (Low) or demanding and complex (High)?"
2. "Physical Demand: How physically demanding was the task? How much physical activity was required? Was the task easy and slack (Low) or demanding and strenuous (High)?"
3. "Temporal Demand: How hurried or rushed was the pace of the task? How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow (Low) or rapid (High)?"

4. "Performance: How successful were you in accomplishing what you were asked to do? How successful were you in performing the task? How satisfied were you with your performance?"
5. "Effort: How hard did you have to work to accomplish your level of performance? How hard did you have to work (mentally and physically) to accomplish your level of performance?"
6. "Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you? How irritated, stressed, and annoyed (High) versus content, relaxed, and complacent (Low) did you feel during the task?" [155]

For this study, the "Raw-TLX" (RTLX) method was chosen[15]. It was proven that, in many cases, RTLX provides as significant data as compared to the original weighted method [24, 57, 154]. Furthermore, as noted by Virtanen et al., the weighting done by 5 out of the 20 pilots in their study was inconsistent [371]. With that in mind, having weighted assessments done by real and naïve pilots would not be appropriate as naïve pilots lacked the knowledge needed in such assessments. Therefore, a Task Load Index score was calculated as follows:

$$\begin{aligned}\text{NASA-TLX} =&((Q_1-1)+(Q_2-1)+(Q_3-1)+((8-Q_4)-1)+\\&+(Q_5-1)+(Q_6-1))*\frac{100}{36},\end{aligned} \tag{4.8}$$

which can be shortened to the form:

$$\text{NASA-TLX} = (Q_1+Q_2+Q_3-Q_4+Q_5+Q_6+2)*\frac{100}{36}. \tag{4.9}$$

In the questionnaire, the scale for questions was 1-7, so to calculate the Task Load Index from 0 to 100, 1 was subtracted from each question. The scale for question 4 (performance) was inverted to not confuse the participant in scoring; otherwise, if the left end of the scale meant "perfect / very high", it would be in contradiction to other questions where the "very high" answers were on the right-hand side. This was reflected in the equation.

---

[15] Used interchangeably as "NASA-TLX" in the rest of this thesis

4.7.6 Objective performance

Recorded flight simulation data included the aircraft's position and orientation, and the forces acting on the aircraft upon landing. These measures were used in performance score calculations (Section 5.3) in order to provide a quantitative assessment of the participant's performance using each inceptor in the trial and to identify any differences among groups.

Apart from HQ, SA and workload questionnaires like CHR, SUS, SART or NASA-TLX, there is no standardised "scoring system" for flight simulation trials. In the past, researchers have used various techniques of assessing pilots' performance, often using MAE [73, 386] and RMS [234, 250, 292, 365], or developing custom equations that suited the experiment design [281]. The lack of standardised measures also occurs in other fields, such as medicine, where, for example, researchers have to develop custom scoring systems in VR training setups [30]. There is no standardised way of assessing performance that combines spatial, temporal, and other measured dimensions that allows easy modification of the weight of each variable. Thus, a new scoring system was developed for the purposes of this trial in order to have a scoring system which is clear and easy to analyse. The equations are presented in Section 5.3. Moreover, this system allowed to compare participants' results, even in the case where one of the tasks was revised at the beginning of the trials - after the first two participants and then after the tenth, as explained in Section 4.5.1.

In order to support the results, considering spatial and temporal results, a performance scoring system and its equations were developed. To date, researchers used either spatial (RMS, MAE) or temporal (time) measures instead. The scoring system presented in this study allowed to consider any measure, and weigh it according to the research needs (similar to NASA-TLX, where the weighting is also present, but for measuring objective performance). Moreover, having a performance score with specified bounds ensured there were no outliers.

4.7.7 Baseline data

In order to measure the performance of the landing tasks, an experienced test pilot was invited to record a baseline landing. The pilot was a 43-year-old professional with 7100 total flight hours in piloting fixed-wing aircraft and 300 total flight hours on engineering or commercial flight simulators, and held a current rating of Saab SF340, with previous ratings of ATR42/72 and BAe Jetstream 31/32. The procedure for recording the baseline landing involved presenting the pilot with a briefing that included the FSS specification, aircraft model description, and task details. After being placed in the cockpit and adjusting the seat, the pilot was allowed to do as many test flights and landing attempts as desired, all of which were recorded. This training took one hour. The scenario setup was an SS-controlled LN task, identical to the one presented to every other participant in this study. Two variants of the landing were performed: an offset landing and a straight landing. After the training and a number of landings (both offset and straight), the pilot chose one of each that, in his opinion, was the best. Those recordings were used as a baseline to compare the performance of other participants in landing scenarios.

It is worth noting that, even with adequate training time, the results showed that the downward G force upon landing was 2.81G for the offset landing and 3.27G for the straight landing. This situation occurred because of how the automation settings were programmed in the experiment design. After reaching the altitude threshold of 50 feet, the autothrottle and spoilers were set to stop the aircraft, which increased the descent rate and, therefore, the G impact on the touchdown.

This page is intentionally left blank.

5

Results and analysis

This chapter shows a detailed presentation of gathered data and an in-depth analysis of the interactions between variables. It also describes the author's findings and recommendations for further study.

5.1 Pilot study results

There are different designs of inceptors applied in the modern flight deck. How do pilots define how to precisely control the aircraft as their intention? Ten professional pilots were asked to take part in the flight simulation trials. They were given tasks described in Section 4.5 and provided feedback using the System usability scale. The aim was to investigate the feasibility of replacing conventional inceptors in aircraft and to assess the experimental design for further study. The results showed that there is a potential for introducing alternate human-computer interaction (HCI) methods in the flight deck, especially in terms of learnability; however, pilots' perceived usability of the alternate controllers was not high. This section summarises the "pilot study" results, published by the author, and shows the potential for further research [197]. The study tested the following hypotheses:

Hypothesis 4 (H_4): *There is a significant difference in pilot's feedback on system usability among three control inputs.*

and

Hypothesis 5 (H_5): *There is a significant difference in pilot's variance of error among three control inputs".*

Preliminary data recorded by the authors before the trials had shown differences in RMS in sidestick and touchscreen controller deflection and performance in keeping the disturbance rejection in the vertical channel (DRV) task's target. Interestingly, sidestick and gamepad results had similar RMS error values, as seen in Tab. 5.1.

Table 5.1: Preliminary results of RMS error values (variance of error) (in degrees) using different inceptors in disturbance rejection vertical DRV task, with the author acting as a pilot.

Inceptor	RMS
Sidestick (SS)	1.61
Gamepad (GP)	1.48
Touchscreen (TS)	2.27

Next, the 'pilot study' trials were carried out. Ten participants holding a valid piloting license, aged from 24 to 63 (mean $M = 39.60$, standard deviation $SD = 12.19$) with fixed-wing total flight hours experience ranging from 800 to 13300 ($M = 5810$, $SD = 3847$) took part in this research[1]. The majority of them held an airliner-type rating (Airbus A320 family or Boeing B737-400/B747-400), and 8 of them were flight instructors. All of them were right-handed.

The results showed that, throughout the experiment, pilots would give the highest SUS scores for the sidestick; however, according to findings of McLellan et al., users tend to put 15-16% higher scores for systems they already know [245]. With that in mind, it could be assumed that the scores for the gamepad and touchscreen would have been higher if participants had previous experience with them. This phenomenon could also be seen in the case of the gamepad - two of the pilots who had previous experience in playing video games using a gamepad (around 50%-75% of time spent playing) had SUS-Total scores much higher than the average: from 71.25 to 92.50 across all four tasks, whereas the mean gamepad's score among all participants was 59.00. Since there was no previous research involving the replacement of the physical sidestick with a touchscreen controller, an assumption could be made that the scores

---

[1] One of the participants completed the trial only using SS and GP, due to their time constraints.

would have been 15% higher if pilots had any experience with piloting the aircraft using the touchscreen technology.

SUS analysis also demonstrated that the DRV task was easiest to perform, while the hardest one was the landing task with disturbance. Boxplots showing the distribution of SUS results can be seen in Fig. 5.1 for usability, Fig. 5.2 for learnability and Fig. 5.3 for total score.[2]

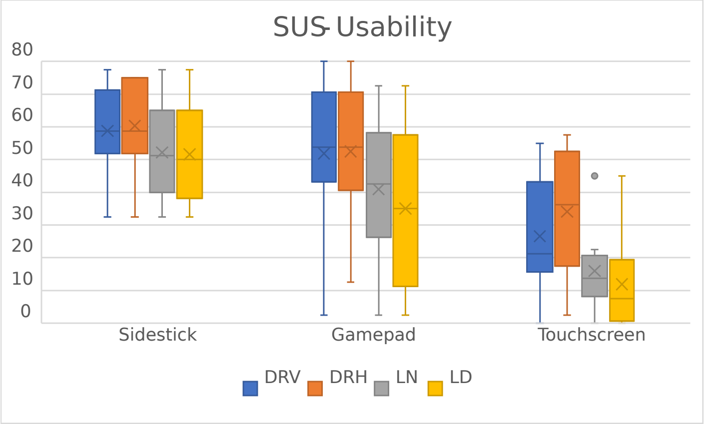


Figure 5.1: System usability scale - usability scores from the pilot study.

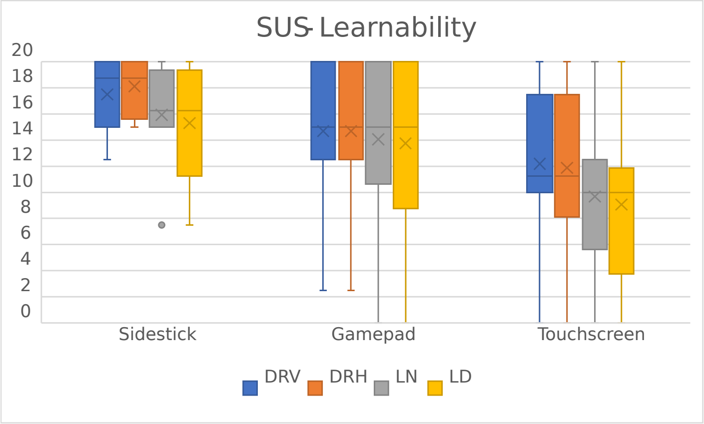


Figure 5.2: System usability scale - learnability scores from the pilot study.

Pilots ranked the sidestick as the highest usability controller, followed by the gamepad and touchscreen. The main reason for this was familiarity - they already have had previous experience with sidestick. Gamepad results were lower than those of

---

[2] The style of plots is different from those in the further analyses due to the fact that the plots in Fig. 5.1, 5.2, and 5.3, along with later Fig. 5.4 were produced right after the pilot study, so several months before trials have finished the main analyses were carried out.

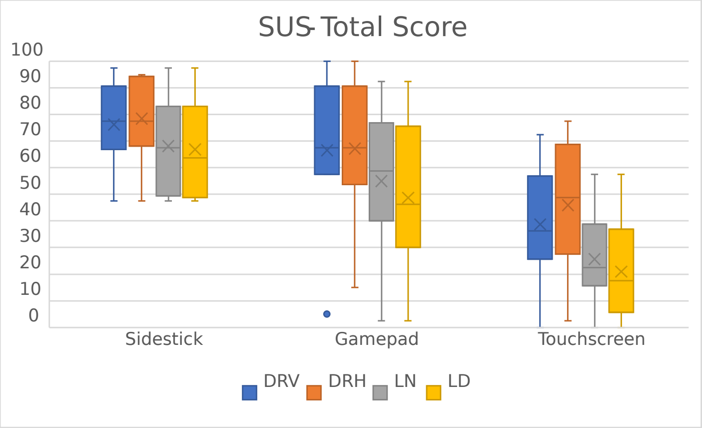


Figure 5.3: System usability scale - total score scores from the pilot study.

sidestick, but there was a much bigger standard deviation - this was because some pilots already experienced using this type of controller in video games (VG). The touchscreen inceptor scored the lowest because none of the participants had experienced this type of controller in the past.

There was also a tendency for significantly higher SUS scores among pilots who indicated an interest in touchscreen technology being introduced in aircraft cockpits - $M = 40.00$ ($SD = 10.51$) for people who liked the idea and $M = 26.46$ ($SD = 23.51$) for people who did not.

It was observed that for the majority of participants, learnability for all three inceptors, especially for disturbance rejection (DR) tasks, was at least satisfactory. Learnability of more challenging landing tasks, especially with the novel touchscreen controller, was somehow lower compared to the rest of the scenarios. Interestingly, the learnability for the gamepad was high, especially for subjects with previous gaming experience.

Randomising the order of the inceptors did not cause a significant change in pilots' behaviour: SUS score for the gamepad was slightly higher (4.5 points on average) than sidestick from participants with sidestick as a first inceptor, however, participants who tested the gamepad and touchscreen first had the same trend for scoring: sidestick > gamepad > touchscreen. An interesting observation could be made here: in DR tasks, the root mean square error (RMS) was lower for the gamepad than sidestick across all the participants, which means that they performed the best using the gamepad (and they did not realise that because it is on the contrary to the majority of the SUS scores). The RMS values can be found in Tab. 5.2, and the distribution of RMS between subjects is shown in Fig. 5.4.

Table 5.2: Root mean square of error (RMS) (deviation from the task objective) averaged across pilots from the pilot study for disturbance rejection vertical (DRV) and disturbance rejection horizontal (DRH). M - mean; SD - standard deviation.

	SS		GP		TS	
	M	SD	M	SD	M	SD
DRV	2.19	1.28	1.41	0.40	4.87	2.18
DRH	2.49	0.98	1.65	0.22	2.88	0.54

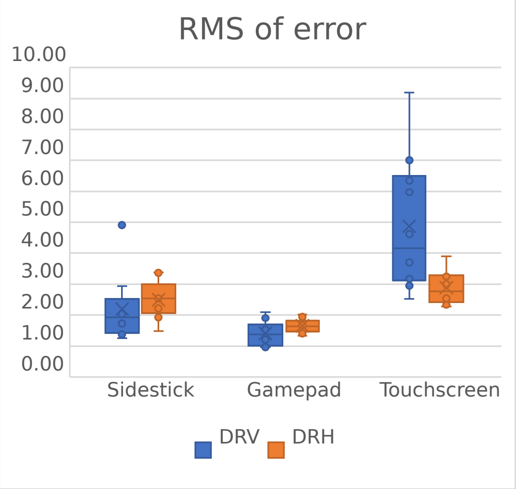


Figure 5.4: Root mean square of error (RMS) (deviation from the task objective) (in degrees) from the pilot study.

An interesting point was raised by one of the participants, who said that landing tasks were easier for them using a gamepad than sidestick, which was reflected in the SUS usability score being higher by 3 points for both landings and their after-trial inceptor preference choice of gamepad. It is worth noting that according to the demographic questionnaire, they have never or hardly ever played VGs or used this type of controller.

Analysis of the results with ANOVA tests showed that the hypothesis (H_4): *There is a significant difference on pilot's feedback on system usability among three control inputs* applied for comparison of sidestick with touchscreen and gamepad with touchscreen, however comparison of SUS scores from sidestick and gamepad did not show significant differences. Hypothesis (H_5): *There is a significant difference on pilot's variance of error among three control inputs* was found to be true when comparing the DRV task results among the three studied inceptors. RMS from the disturbance rejection in the horizontal channel (DRH) task has shown that the gamepad was a more accurate inceptor than the rest (RMS Mean equal to 2.49, 1.65 and 2.88, respectively for sidestick, gamepad and touchscreen). More detailed results can be found in the author's article [197].

The purpose of the pilot study was to initially investigate pilots' behaviour using three different inceptors. The two formed hypotheses (H_4) and (H_5) were checked against the results and confirmed to be true in most cases. While there is still a lot of work in touchscreen technology in flight decks and there is a long way before replacing the physical inceptors with digital ones, this research showed that even though the touchscreen controller scored the lowest on a SUS, the majority of pilots were able to put the aircraft on the ground in these challenging circumstances. As this was only a 'pilot study', the sample size was small, and conclusions made based on the experimental results were mostly applicable to airline pilots who have already had training and experience with conventional inceptors. This could have introduced bias in the statistical metrics because they lacked the same training on the other alternative inceptors. Further trials, with results presented in the following sections of this chapter, included a comparison of pilots with non-pilots (”naïve pilots”) to see the effect of learnability: non-pilots with no experience using the sidestick were not biased by the habits learned in flying schools. Furthermore, the results were investigated using statistical analyses.

5.2 Mean absolute error and root mean square error

Initially, participant's variance from a target was calculated using two techniques - mean absolute error (MAE) [73, 386], and root mean square error (RMS[3]) [234, 250, 292, 365]. Both methods are commonly used to measure the accuracy of continuous variables [400]. Both methods are negatively-oriented scores, so lower MAE and RMS values mean better performance, but they can give different results on the same set of data. MAE is less penalising for large errors (outliers), whereas RMS is more sensitive to them (gives larger value in overall results), which gave more margin if the pilot was only slightly off the target [182]. The RMS will always be equal to or bigger than MAE. RMS is usually preferred in spatial measures [283]. For the purpose of this study, both methods were calculated to check the correlation between them, presented in Section 5.2. The equations for MAE and RMS are:

---

[3] In this study, the abbreviation ”RMS” is used instead of ”RMSE”; they both concurrently mean ”Root Mean Square Error”.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_{T_i} - x_i|, \tag{5.1}$$

and

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_{T_i} - x_i)^2}, \tag{5.2}$$

where n is the number of samples (time frames), x_{T_i} is the target value for the frame i, and x_i is the measured value in the frame i. Both methods measure the average magnitude of the error of the estimates and are expressed in the same units as the measured and target values.

It was assumed that the RMS and MAE values in each scenario (DRV, disturbance rejection in vertical channel - DRH, landing with no disturbance - LN, and landing with disturbance - LD) would have a significant positive relationship between each other. Bivariate Pearson Correlation with one-tailed test of significance was performed [171][4]. The results are shown in Tab. 5.3. A highly significant value of $R > 0.9$ ($p < .001$) in every scenario confirmed that there was a near-perfect positive correlation, so the assumed relation was confirmed. Therefore, for the rest of the analysis, RMS was chosen, because it takes larger errors into consideration and was found to be mostly used in aviation research [233, 250, 292, 365].

Table 5.3: Results of Bivariate Pearson Correlation r with one-tailed test of significance p to investigate the correlation between MAE and RMS values.

Scenario	r	p
DRV	0.993	$< .001$
DRH	0.990	$< .001$
LN	0.989	$< .001$
LD	0.977	$< .001$

---

[4] One-tailed test of significance was chosen because the direction of association was clearly visible in advance

5.3 Performance Score

For the purpose of this study, a scoring system for participants' performance was developed for each type of scenario, marked as the "Performance Score" (PS). The score equations for DR (PS_{DR}) and landing (PS_L) scenarios, respectively Eq. 5.4 and Eq. 5.8, were explained in this section. The Performance Score PS is a unitless measure, ranging from 0 to 100, where 100 means the best performance. It is calculated by subtracting "Penalty Points" PP from 100. The DR scenario score PS_{DR} depends on spatial and temporal measures (respectively RMS of deflection and total time outside the acceptable and desirable margins, given in the form of a certain amount of PP), while the landing scenario Performance Score PS_L depends only on spatial measures (the flight path, touchdown point and downward G deviance from the baseline landing, also translated into PP). The equations can be easily adjusted to suit different experimental designs by changing measured variables and their weights. Notation $Norm$, defined as:

$$Norm(L \leq x \leq U) = \frac{Clamp(L \leq x \leq U)}{U - L} = \frac{min(max(x - L, L), U - L)}{U - L}, \tag{5.3}$$

is used for normalising the given x value in equations for penalty points PP, clamped between lower L and upper U bounds. The resulting value of $Norm$ is always in the range of $0 - 1$, hence each penalty point equation is multiplied by a weight value. The total number of penalty points PP for each Score equation PS can not exceed 100.

5.3.1 Performance Score for disturbance rejection tasks (DRV and DRH)

The equation for calculating the PS for DR tasks is:

$$PS_{DR} = 100 - (PP_{AccRMS} + PP_{DesRMS}) - PP_T, \tag{5.4}$$

where:

- PS_{DR} - Performance Score for disturbance rejection tasks (DRV and DRH),
- PP_{AccRMS} - Penalty points for deviance from the acceptable margin (spatial measure) - Eq. 5.5,
- PP_{DesRMS} - Penalty points for deviance from the desirable margin (spatial measure) - Eq. 5.6, and

- PP_T - Penalty points for time outside acceptable and desirable margins (temporal measure) - Eq. 5.7;

$$PP_{AccRMS} = Norm(0 \leq E_{AccRMS} \leq U_{Acc}) * 25, \tag{5.5}$$

where:

- E_{AccRMS} - RMS outside acceptable margin, clamped in a range of $R_{Acc} \in \langle L_{Acc}; U_{Acc} \rangle$, where:
  - R_{Acc} - Range for E_{AccRMS}, and
  - L_{Acc}, U_{Acc} - Lower and Upper bounds of R_{Acc}. For DRV: see Tab. 5.4. For DRH: $R_{Acc} \in \langle 0; 3 \rangle$ [deg];

$$PP_{DesRMS} = Norm(0 \leq E_{DesRMS} \leq U_{Des}) * 25, \tag{5.6}$$

where:

- E_{DesRMS} - RMS outside desirable margin, clamped in a range of $R_{Des} \in \langle L_{Des}; U_{Des} \rangle$, where:
  - R_{Des} - Range for E_{DesRMS}, and
  - L_{Des}, U_{Des} - Lower and Upper bound of R_{Des}. For DRV: see Tab. 5.4. For DRH: $R_{Des} \in \langle 0; 8 \rangle$ [deg];

$$PP_T = \frac{T_{outAcc} + T_{outDes}}{T_{total} * 2} * 50, \tag{5.7}$$

where:

- T_{outAcc} - time outside acceptable margin,
- T_{outDes} - time outside desirable margin, and
- T_{total} - total time of the task.

Table 5.4: Ranges R_{Acc} and R_{Des} of errors E_{AccRMS} and E_{DesRMS} in DRV scenario, showing Lower and Upper bounds L_{Acc}, U_{Acc} and L_{Des}, U_{Des}: $R_{Acc} \in \langle L_{Acc}; U_{Acc} \rangle$, $R_{Des} \in \langle L_{Des}; U_{Des} \rangle$

Participants	R_{Acc}	R_{Des}	Units[5]
1-2	$\langle 0; 2.5 \rangle$	$\langle 0; 4.5 \rangle$	deg
3-10	$\langle 0; 4 \rangle$	$\langle 0; 5 \rangle$	deg
11+	$\langle 0; 1700 \rangle$	$\langle 0; 1700 \rangle$	ft

---

[5] Although RMS value is unitless, this column indicates units of the measured and target values.

In PS_{DR} (Eq. 5.4), the spatial $PP_{AccRMS} + PP_{DesRMS}$ and temporal PP_T Penalty Points are being equally weighted: for spatial and temporal, maximum penalty points are 50 each. The weight was chosen after correlating the spatial and temporal results, shown in Tab. 5.5. For DRV scenario, results were moderately correlated ($r \geq .583$, $p < .001$), and for DRH they were highly correlated ($r \geq .858$, $p < .001$). In PP_{AccRMS} and PP_{DesRMS}, RMS errors E_{AccRMS} and E_{DesRMS} are equally weighted, and E_{DesRMS} includes values outside acceptable margin. Values of deflection and time outside acceptable margin are effectively "penalised" twice, highlighting the difference between desirable and acceptable margins in Penalty Points. This also applies to temporal Penalty Points, which are calculated by summing the proportion of time spent outside the acceptable and desirable margins, T_{outAcc} and T_{outDes}, to total task time T_{total}. The RMS error of Each spatial error has a specified range, which means if the measured value was lower or higher than the range's bounds, the penalty score was set to a zero or maximum weight value, respectively. The ranges were chosen based on overall trial results - DR error bounds were assumed by taking a value between the maximum and average RMS error of all participants, $(Max_E + Avg_E)/2$, for each scenario respectively. Differences in ranges between participants for DRV, seen in Tab. 5.4, occur due to revisions of task definition during the early trials.

Table 5.5: Results of Two-tailed Bivariate Pearson Correlation r between the RMS error E_{RMS} and time outside T_{out} desirable and acceptable margins for DR scenarios. N - number of samples.

E_{RMS} x T_{out}	Margin	r	p	N
DRV	Desirable	.583	$< .001$	221
	Acceptable	.644	$< .001$	221
DRH	Desirable	.891	$< .001$	221
	Acceptable	.858	$< .001$	221

5.3.2 Performance Score for landing tasks (LN and LD)

The PS for LN and LD scenarios is calculated as follows:

$$PS_L = 100 - PP_{RMS} - (PP_{AccTD} + PP_{DesTD}) - PP_G, \tag{5.8}$$

where:

- PS_L – Performance Score for landing tasks (LN and LD),
- PP_{RMS} - Penalty points for deviance from baseline landing's flight path - Eq. 5.9,
- PP_{AccTD} - Penalty points for acceptable touchdown location error from baseline landing - Eq. 5.10,
- PP_{DesTD} - Penalty points for desirable touchdown location error from baseline landing - Eq. 5.11, and
- PP_G - Penalty points for maximum downwards G compared to baseline landing - Eq. 5.12.

(if the participant did not manage to land, maximum penalty points of PP_{AccTD}, PP_{DesTD} and PP_G were taken out of PS_L);

$$PP_{RMS} = Norm(L_{RMS} \leq E_{RMS} \leq U_{RMS}) * 75, \tag{5.9}$$

where:

- E_{RMS} - RMS from baseline landing's flight path, clamped in a range of $R_{Acc} \in \langle L_{RMS}; U_{RMS} \rangle$, where:
  - R_{RMS} - Range for E_{RMS}, and
  - L_{RMS}, U_{RMS} - Lower and Upper bound of R_{RMS}. $R_{Acc} \in \langle 35; 1000 \rangle$ [m];

$$PP_{AccTD} = Norm(L_{AccTD} \leq E_{AccTD} \leq U_{AccTD}) * 5, \tag{5.10}$$

- E_{AccTD} - Distance outside the acceptable margin of baseline landing's touchdown location, clamped in a range of $R_{AccTD} \in \langle L_{AccTD}; U_{AccTD} \rangle$, where:
  - R_{AccTD} - Range for E_{AccTD}, and
  - L_{AccTD}, U_{AccTD} - Lower and Upper bound of R_{AccTD}. $R_{AccTD} \in \langle 152; 1000 \rangle$ [m];

$$PP_{DesTD} = Norm(L_{DesTD} \leq E_{DesTD} \leq U_{DesTD}) * 5, \tag{5.11}$$

- E_{DesTD} - Distance outside the desirable margin of baseline landing's touchdown location, clamped in a range of $R_{DesTD} \in \langle L_{DesTD}; U_{DesTD} \rangle$, where:
  - R_{DesTD} - Range for E_{DesTD}, and
  - L_{DesTD}, U_{DesTD} - Lower and Upper bound of R_{DesTD}. $R_{DesTD} \in \langle 76; 500 \rangle$ [m];

$$PP_G = Norm(L_G \leq G_{TD} \leq U_G) * 15, \tag{5.12}$$

where:

- E_G - Difference in maximum downwards G from baseline landing, clamped in a range of $R_G \in \langle L_G; U_G \rangle$, where:
  - R_G - Range for E_G, and
  - L_G, U_G - Lower and Upper bound of R_G. $R_G \in \langle 3.3; 4.5 \rangle$ [G].

Since the experiment's focus was on flight path control, the deflection from baseline flight path E_{RMS} was the most important aspect in PS_L (Eq. 5.8), attributing to 75% of the Score. The remaining 25% of PS_L is shared between the touchdown distance error outside the acceptable/desirable margin and maximum downwards G upon touchdown. This weighting allows participants with diminutive RMS flight path error to still achieve a relatively high Score, regardless if they managed to land or not. This is reflected in the results, where the maximum and average score for participants who did not land is: $Max = 70.98$, $Avg = 51.23$ ($SD = 21.66$, $N = 40$) in LN scenario and $Max = 70.49$, $Avg = 44.71$ ($SD = 22.97$, $N = 51$) in LD scenario. In another aspect, the similarity of the scores shows the learning curve, as the LD scenario was more challenging. PP_{AccTD} and PP_{DesTD} are equally weighted, so having the landing distance error outside the acceptable margin is penalised double (similarly to DR's PP_{AccRMS} and PP_{DesRMS}). PP_{AccTD} and PP_{DesTD} have a weight of 10, while PP_G was assumed to have a weight of 15, in order to recognise the higher importance of landing smoothly than in the right touchdown zone. Table 5.6 shows that there is only one significant correlation between E_{RMS}, E_{TD}, and E_G, which is between the first two ($r = .850$, $p < .001$ for LN, and $r = .592$, $p < .001$ for LD; the correlation has been checked without taking margins into account). This means that there is a significant correlation between the flight path and the landing location, however good results in those aspects do not mean that the participant landed smoothly (or "did not guarantee a smooth landing"?). Error range for landing flight path was assumed by summing the average and SD of all participants' RMS error ($Avg_E + SD_E = 339 + 620 = 959$ for LN and $Avg_E + SD_E = 438 + 759 = 1197$ for LD), then averaging the result and rounding it down to 1000. The lower bound was assumed to be an average of minimum RMS error values of LN and LD ($Min_{LN} = 42$, $Min_{LD} = 27$), rounded to 35. Lower bound ranges for touchdown location were based on Mitchell's definition of desirable and adequate (acceptable) touchdown aimpoint margins[6] [254], and upper ranges were assumed by summing the average and SD of touchdown error values across all participants ($Avg_{TD} + SD_{TD} = 488 + 692 = 1180$ for LN and $Avg_{TD} + SD_{TD} = 487 + 672 = 1159$ for LD), then averaging the result and

---

[6] Touchdown within +/- 250 feet of aimpoint for desired performance; touchdown within +/- 500 feet of aimpoint for adequate performance.

rounding it down to 1000 (in meters). For the desirable lower bound, this value was divided by 2. Any values exceeding the range were treated as a maximum error. For G_{TD}, if the participant landed with less G impact than the baseline pilot (3.3G), he received no penalty points PP_G. The unusually high value of the G for the baseline pilot is explained in Section 4.7.7. The value of 4.5G was assumed as a "crash" threshold [122]. It was assumed that, under normal conditions, the baseline pilot would not exceed 1.5G upon landing. In Airbus A320, a landing with over 2.6G is considered to cause structural damage [392]. Thus, the upper margin was taken by subtracting the 1.5G from the A320's structural damage threshold value, and then adding the baseline pilot's downwards G value and rounding it to the resolution of 0.5: $U_G = 2.6 - 1.5 + 3.3 = 4.4 \approx 4.5$. Another limitation in this was the fact that the trials were conducted before the upgrade of the outside visual PC. Although all the participants and the baseline pilot had flown using the same hardware setup (so all of them had been subjected to a control lag introduced due to the decreased frame rate), so the error due to this was not a between-subject factor; it is, however, one of the factors that explain the unrealistically high downwards G upon landings, even within experienced pilots, and justifies the high G value set as a lower bound in the PS Landing equation.

The temporal aspect in PS_L was insignificant, as the aircraft's speed was handled by the Automatic Flight System, and the desirable/acceptable margins were not specified for the Landing scenarios. Because of this, the time to the touchdown was similar for every participant.

Table 5.6: Results of Two-tailed Bivariate Pearson Correlation r between the RMS error E_{RMS}, touchdown location error E_{TD} and maximum downwards G E_G for Landing scenarios. **Bold** values mean that the correlation was significant. N - number of samples

Scenario	Variables	r	p	N
LN	E_{RMS} x E_{TD}	**.850**	$<$ .001	181
	E_{RMS} x E_G	.138	0.64	181
	E_{TD} x E_G	.121	.105	181
LD	E_{RMS} x E_{TD}	**.592**	$<$ .001	170
	E_{RMS} x E_G	-.011	.890	170
	E_{TD} x E_G	-.045	.558	170

5.4 Participants grouping

In order to see the effect of between-subject factors on the results, participants were assigned to different groups based on their flying experience (either real aircraft or flight simulators) and video game experience.

Regarding the flying experience, participants were divided into 3 groups:

A High-experienced pilots (fixed wing total hours ≥ 250[7] or flight simulator total flight hours ≥ 500).
B Low-experienced pilots (fixed wing total hours between ≥ 10[8] and < 250 OR flight simulators between ≥ 50 and < 500)
C naïve pilots (fixed wing total hours < 10 and flight simulators < 50)

Regarding the video games experience, participants were divided as follows:

A High-experienced gamers (playing or used to play video games for more than 3 hours per week)
B Low-experienced gamers (playing or used to play video games for less than 3 hours per week)
C Non-gamers (never or hardly ever played video games)

The groups defined here were used in further analyses, although, in some cases, the gaming frequency was also investigated.

5.5 Scenario correlations and hierarchical clustering

This study collected a large amount of data, and analyses of multiple factors are prone to various errors [124]. Therefore, the identification and elimination of redundant and non-significant measures had to be made to minimise those errors. This section, along with Sections 5.5.2, 5.6 and 5.6.1 addressed that concern.

Agglomerative hierarchical clustering was performed to investigate if the results

---

[7] There have been some studies that indicated that just the total flight hours is not an adequate measure of a pilot's experience; however, those studies were focusing on accident and incident investigation [364]. Here this value was assumed based on other publications [176, 354] and the uniform distribution of participants.

[8] 10 hours was chosen as a margin because some participants have responded with such low values in the "fixed-wing total flight hours" field and did not indicate holding any piloting license

of each subjective rating (Cooper-Harper Rating Scale - CHR, system usability scale - SUS and its subdimensions, situational awareness rating technique - SART and its subdimensions, and Task Load Index - NASA-TLX) could be clustered and correlated. For this purpose, MATLAB code was prepared, which executed the following steps:

1. Invert the scales in which higher values mean worse performance/experience (CHR, SART-D and NASA-TLX) by subtracting the given score from a sum of the maximum and minimum achievable score: $x_{inverted} = (x_{max} + x_{min}) - x$, where x is the corresponding score for each participant.
2. Create the distance matrix with all questionnaire results as an input using *pdist* function. Jaccard coefficient method was used as a distance metric, as it gives a normalised distance between two samples, measured between 0 and 1.
3. Group the factors by linking pairs in close proximity using *linkage* function with a "Ward" method, which computes the inner squared distance between the clusters using the minimum variance algorithm. This method also tends to find "natural" clusters better than other methods [164]. Although some information can be found that this method is appropriate for Euclidean distances only [241], evidence in the literature exists that this method is applicable with Jaccard Distance Measure, and in some cases, it produces better results than with Euclidean distances [120, 146, 164, 376]. Other available distance and linkage methods were investigated as well, however, the combination of Jaccard and Ward produced the best results. This created a binary hierarchical cluster tree with 9 distinct clusters.
4. Verify the dissimilarity - calculate the cophenetic correlation coefficient with *cophenet* function. Resulting value close to 1 meant that the clustering solution reflects the data accurately.
5. Divide the hierarchical tree into clusters using *cluster* function and visualise the results on a dendrogram plot.

The resulting dendrogram plot is shown in Fig. 5.5 (top). Each colour on the plot represents a different cluster. Cophenetic correlation coefficient $c = .769$ verified the dissimilarity of the linkage operation. The result clearly shows that there were strong similarities between the results within each score type. The results were verified by calculating their correlations and plotting them on the graph in Fig. 5.6. This finding allowed the calculation of the averaged result within each scale, as presented in the following equation:

$$x_{avg} = Avg(x_{DRV}, x_{DRH}, x_{LN}, x_{LD}), \tag{5.13}$$

where x is the participant's subjective score from each of the scales from the questionnaire: CHR, SUS (SUS-U, SUS-L, SUS-Total), SART (SART-S, SART-U, SART-D, SART-Total), and NASA-TLX.

Dependent variables could be split into two "families": all the subjective measures were calculated in the same way between the scenarios, while the PS had a different way of calculation for DRV, DRH and landing scenarios. Because of this, hierarchical clustering for PS was performed separately. The resulting dendrogram in Fig. 5.5 (bottom) shows that the distance between the scenarios was high compared to the subjective measures (top). Therefore, the decision was to combine subjective measures from each scenario together, while leaving the objective measures separate for objective measure (PS).

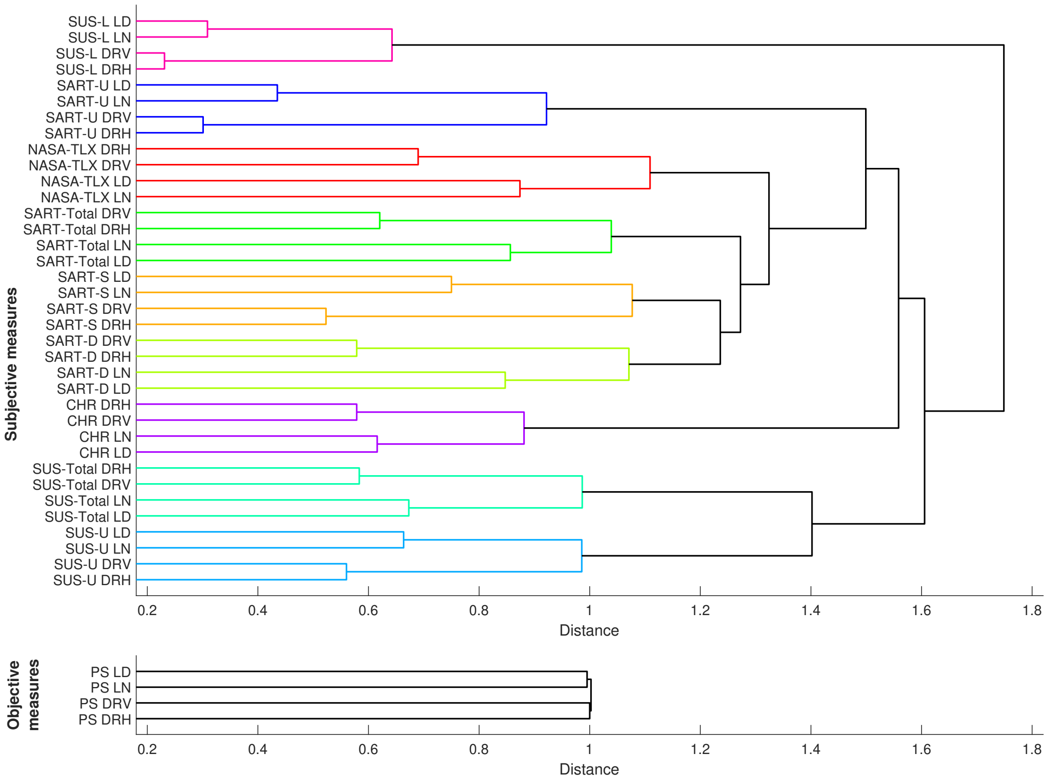


Figure 5.5: Dendrograms of agglomerative hierarchical clustering of the subjective (top) and objective (bottom) measures between the scenarios. Each colour different than black indicates a viable cluster.

To further confirm that the averaged scores correlated with each scenario's results,

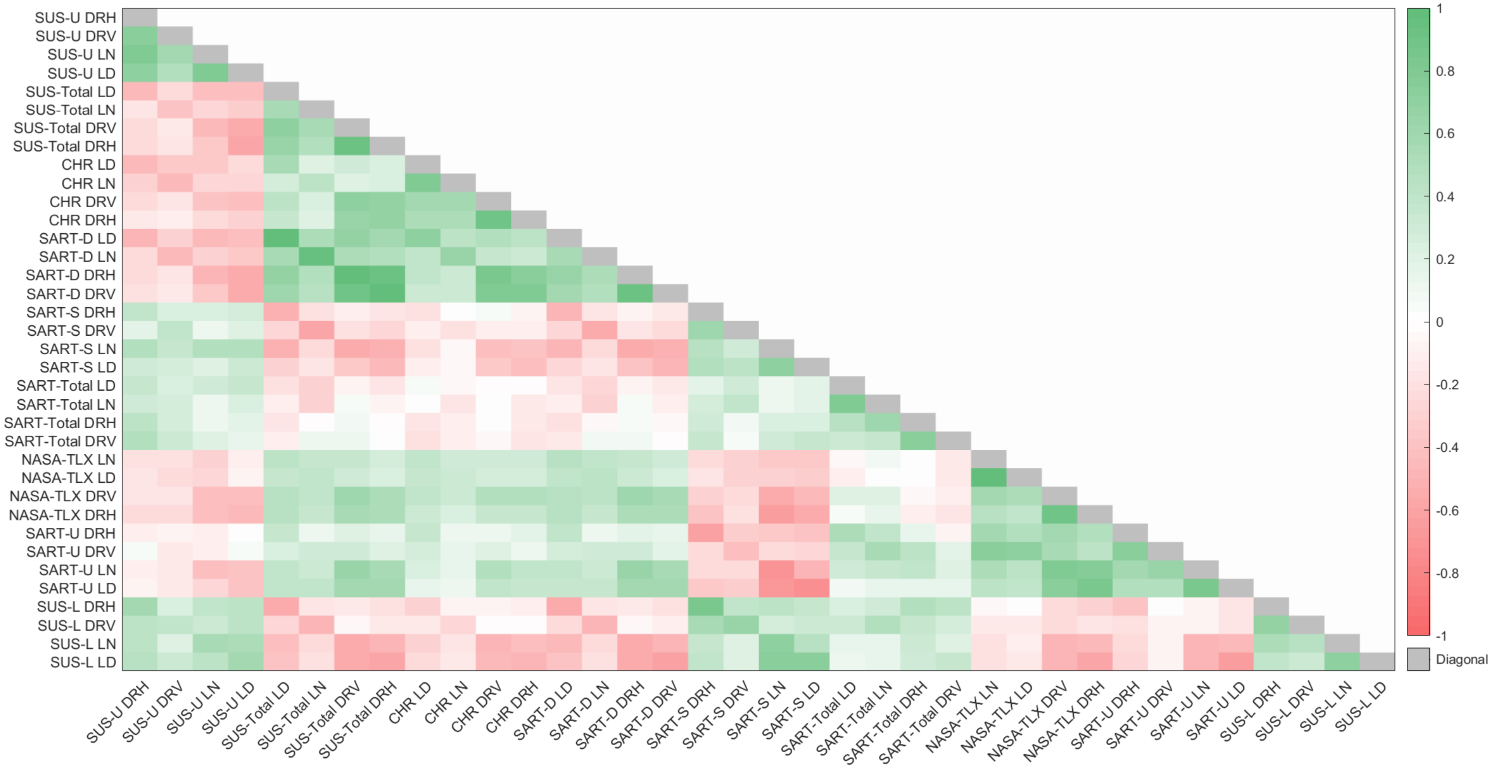


Figure 5.6: Heatmap presenting correlations between subjective measure factors. The factors are ordered based on hierarchical clustering. Along the diagonal line, distinct clusters between the scenarios for each factor can be seen. MATLAB function *multigradient* was used to create a custom gradient [202].

Bivariate Pearson Correlation R with two-tailed test of significance p was performed, based on the hierarchical clustering implemented before. Detailed results can be found in Tab. A.1 in the appendix. An average value of $R = .855$ and constant significance level $p < .001$ across all results shows that there was a strong correlation between the averaged result and the result of each scenario.

Another finding from the clustering result in Fig. 5.5 was that SART-S and SART-D had the strongest connection (the shortest distance where they could be joined as a new cluster), so the results were closely correlated. This happens when the measured system's supply and demand have similar scores; in this case, it indicated that all three inceptors were doing exactly what the users were expecting them to do, which was controlling the aircraft. Moreover, this analysis revealed that the scenario pairs DRV/DRH and LN/LD were also similar, independent of the grouping method.

5.5.1 Kendall's coefficient of concordance

In order to investigate the concordance between participants in the averaged rating scale and to support the inclusion or exclusion of a grouping variable in analyses (Section 5.6), Kendall's coefficient of concordance W was calculated. Kendall's coefficient of concordance is a measure of how well the ranks of two or more groups of data (repeated measures) "agree" with each other. To compute this, another Matlab function was created. The input was a matrix of the scores for each inceptor for the rating scale in question. The data set was first checked for any invalid entries (such as missing values or "NaN" - "Not a Number"), and Kendall's coefficient was calculated using a modified function, which is available online [170]. Resulting values of W ranged from $W = .172$[9] for SART-U and SART-Total (slight agreement), up to $W = .527$ for SUS-Total (moderate agreement). The results for all factors are presented in Tab. 5.7. None of the factors showed a substantial agreement between the inceptors scoring. This means that there was no redundant measure factor when comparing the three inceptors at this stage.

---

[9] There was a lower value, $W = .038$ for SART-S, although the result was not significant ($p = .064$)

Table 5.7: Results of Kendall's coefficient of concordance W calculations for every factor averaged from all four scenarios.

Factor (averaged)	W	χ^2	p
CHR	.326	47.589	$< .001$
SUS-U	.500	73.068	$< .001$
SUS-L	.178	26.000	$< .001$
SUS-Total	.527	77.014	$< .001$
SART-D	.226	32.986	$< .001$
SART-S	.038	5.507	.064
SART-U	.172	25.178	$< .001$
SART-Total	.172	25.178	$< .001$
NASA-TLX	.350	50.361	$< .001$

5.5.2 Performance Score correlation between the scenarios

Following the subjective measure analyses, Bivariate Pearson Correlation r with two-tailed test of significance p was performed for PSs in each scenario within a controller to check if they could also be correlated between scenarios. It has to be pointed out that the scores for DRV, DRH and LN/LD scenarios were calculated differently, so the results were not expected to be as correlated as those of subjective scales. Results have shown that the correlations vary between $r = .375$ and $r = .736$ ($p < .001$), but there was no visible pattern.

5.6 Initial factor analyses

This study analysed a large number of input and output variables. The demographic data gained from each participant before the test was to distinguish several grouping factors (independent variables - "groups"). During the study, measured factors (dependent variables - "factors") were recorded for each inceptor and scenario. Investigation in Section 5.5 showed that the subjective measures between the scenarios were highly correlated, so the average score for each factor was calculated. In order to find which groups significantly influence which factors, two-way repeated measures ANOVA (rANOVA) was performed for each combination of Group and Factor (9 x 13), resulting in 117 sets (rows) of data. This was done by creating a MATLAB script which ran all the analyses and saved them to a file. Each row contained results of Mauchly's test, epsilon adjustment, and results of rANOVA with effects of group, inceptor, and interaction between them. Additionally, each row contained flags that indicate if the group, inceptor type or interaction were potentially impacting the results (with a 20% tolerance threshold). Such high tolerance was assumed to find trends among the factors and to minimise false negative results (effects that had a significant effect on the measure but would be marked as non-significant in this analysis). At this stage, multiple testing correction was not applied. Instead, Bonferroni and Tukey HSD[10] corrections, along with analytical expertise, were employed in later ANOVA tests to eliminate potential false positive outcomes. Kendall's coefficient of concordance test results, assigned to each group, were also added to the results to support the findings (see Section 5.5.1).

Participants were grouped by: gender, handedness, attitude towards touchscreens in flight deck (TS attitude), gaming frequency (VG frequency; before assigning to VG groups), gamepad (GP) usage, mobile games (MG) usage, inceptor order, flight experience (FE) group, and video game (VG) group. The factors were: CHR, SUS-U, SUS-L, SUS-Total, SART-D, SART-S, SART-U, SART-Total, NASA-TLX, and PS. Subjective factors were averaged between scenarios, while the objective measure (PS) was separate for each task. Mauchly's test was non-significant in all cases, so the sphericity of data was not violated.

It was assumed that the groups that would show significance in these analyses (5% tolerance level)[11] had the chance to have a potential main effect on the corresponding dependent variable, while interactions that were not significant but showed a certain trend (5%-20% tolerance level)[12], if not false negative, had a possibility to be a confounding or adjustment factor for that dependent variable or are a "hidden" factor in another group. The visual representation of results with $p < .2$ can be found in Tab. 5.8, and exact data output can be found in Tab. A.2 in the appendix.

There were eight significant results ($p < .05$ or less). Groups with significant effects were gender (on 2 factors), TS attitude (1 factor), and FE group (5 factors). Grouping by handedness, VG frequency, GP usage, MG usage, inceptor order, and VG group did not have a significant effect on any of the results.

---

[10] *Honestly Significant Difference.*

[11] 5% tolerance level is known as "type I error", or the "α-level", and the margin for significance is $p < .05$. It indicates the probability of false outcome [124].

[12] 20% tolerance level is known as "type II error", or the "β-level", and its margin is $p < .2$ [79].

On the other hand, there were 22 results with a potential trend (between $p \geq .05$ and $p < .2$). It indicated that the groups: gender (on 4 factors), handedness (2 factors), TS attitude (1 factor), VG frequency (1 factor), GP usage (2 factors), MG usage (5 factors), inceptor order (2 factors), FE group (1 factor), and VG group (4 factors) were worth investigating further (within those factors), as they might have been falsely categorised as non-significant, or were a hidden factor within another group.

From those results, the following conclusions were formed (with an assumption that there was any effect of the averaged subjective score, or more than two effects of PS scenarios on each factor):

- Gender potentially had a trending effect on the learnability aspect (SUS-L) and overall usability score (SUS-Total), as well as performance (PS) (significance in DRV and LD, and trend in DRH and LN);
- TS attitude - significance on usability scores (SUS-U) and a trend on overall usability (SUS-Total);
- GP usage - only trend on the experienced SA demand (SART-D);
- MG usage - trend on learnability (SUS-L), SA demand (SART-D), and PS (DRV, LN and LD);
- FE group - significance on usability (SUS-L) and PS (all scenarios), and a trend on overall usability (SUS-Total);
- VG Group - a trend on usability (SUS-U), learnability (SUS-L) and overall usability (SUS-Total), and perceived workload (NASA-TLX).

From all of those groups, FE group had the most significant effects, which was convergent with the initial assumption that there would be differences between real and naïve pilots. Moreover, pilots might have a higher mental capacity to adapt to a new inceptor, as their previous experience in piloting an aircraft would let them focus more on a novel aspect of the experiment. On the other hand, it only affected the perceived usability of the controller and the direct performance in flight. This might be an indication that pilots' SA and perceived workload did not differ considerably from non-pilots.

Table 5.8: Effects of grouping factors (independent variables) on subjective scores (averaged) and PS. Significance ($p < .05$) is marked with ✓, while a trend ($.05 \leq p < .2$) is marked with ∼. Blank cell means that significance was $p > .2$. Exact values can be found in Tab. A.2 in the appendix.

Factor	Gender	Handedness	TS attitude	VG freq.	GP usage	MG usage	Inceptor order	FE group	VG group
CHR									
SUS-U			✓					✓	∼
SUS-L	∼					∼			∼
SUS-Total	∼		∼					∼	∼
SART-D					∼	∼			
SART-S									
SART-U									
SART-Total									
NASA-TLX									∼
PS DRV	✓	∼		∼		∼	∼	✓	
PS DRH	∼							✓	
PS LN	∼	∼			∼	∼	∼	✓	
PS LD	✓					∼		✓	

5.6.1 Age correlation

Age can potentially impact an individual's understanding and performance of complex systems, such as those found in aircraft. As people age, they may experience a decline in cognitive abilities, such as memory and problem-solving skills, which could affect their ability to understand and operate complex systems [93]. On the other hand, many older individuals maintain their cognitive abilities and physical capabilities well into old age and might be just as capable of understanding and operating complex systems as younger individuals. In an example study, professional pilots, having years of flight experience, were more likely to perform better in aircraft-related tasks [189]. To further investigate the effects of demographic characteristics on the results, this section answers if the age of the participants was a significant between-subjects factor in some of the measures. SART-U and PS scores were investigated first, as understanding and performance in using complex systems are commonly linked to ageing [242, 346]. Since age is a numerical variable, it would be difficult to perform a rANOVA. Therefore, the Bivariate Pearson Correlation R with one-tailed test of significance[13] significance was performed to check if a correlation between age and SART-U and PS exists. Detailed results, with exact R and p values, can be found in Tab. A.3 in the appendix. The results showed that there was a slight trending (albeit non-significant) negative correlation between age and understanding for the three inceptors. If the tolerance was elevated to 10%, SS and GP showed a trending small negative correlation for SS and GP ($p < .1$)[14]; the coefficient was not significant for TS. Interestingly, the results indicated that there was a significant small to medium negative correlation between the age and the results using GP (in DRH and LD scenarios, $p > .05$), and TS (in DRV, $p < .001$, DRH, $p < .01$, LN, $p < .01$, and LD, $p < .1$[15] scenarios). This signified that in those cases, younger participants had a slightly better understanding of the inceptors' usage and performed better using the TS controller, which was convergent with observations by McClumpha et al. that younger pilots tend to have a better understanding of the system they use [242].

Those initial analyses indicated that age might be an important factor in introducing new inceptors in the flight deck. Therefore, further analyses were performed, where all dependent variables from the study were examined. Moreover, the results of each scenario were inspected separately to further support the hierarchical clustering of the subjective measures from Section 5.5. The results, with exact R and p values, can be found in Tab. A.4 in the appendix.
The correlations for SS were:

- CHR - potential small positive correlation between averaged scenarios (avg, $p < .1$[16]), with further analysis for DRH ($p < .05$), and LN ($p < .1$)[16];
- SUS-L - small positive correlation between averaged scenarios (avg, $p < .05$), with LN ($p < .05$), and LD ($p < .01$);
- NASA-TLX - small positive correlation between averaged scenarios (avg, $p < .05$), with DRV ($p < .05$), DRH (medium positive correlation, $p < .01$), and potentially LN ($p < .1$)[16];

---

[13] One-tailed test was chosen as there was no indication whether there would be a correlation.

[14] According to Cohen, the absolute value of the coefficient $0.1 \leq R < 0.3$ means that there is a small correlation, $0.3 \leq R < 0.5$ indicates a medium correlation, and $R \geq 0.5$ is classified as large correlation [80].

[15] If the tolerance was elevated to 10%.

[16] Tolerance elevated to 10%, showing a potential trend.

and for GP were:

- CHR - small positive correlation between averaged scenarios (avg, $p < .05$), with DRV ($p < .05$), DRH ($p < .05$), and potentially LN ($p < .1$)[16];
- SUS-U - small negative correlation between averaged scenarios (avg, $p < .01$), with DRV (medium negative correlation, $p < .01$), DRH (medium negative correlation, $p < .001$), and LN ($p < .05$);
- SUS-L - potential small negative correlation between averaged scenarios (avg, $p < .1$)[16], with DRV (medium negative correlation, $p < .01$), DRH (medium negative correlation, $p < .01$), and potentially LN ($p < .1$)[16];
- SUS-Total - small negative correlation between averaged scenarios (avg, $p < .01$), with DRV (medium negative correlation, $p < .001$), DRH (medium negative correlation, $p < .001$), and LN ($p < .05$);
- SART-D - medium positive correlation between averaged scenarios (avg, $p < .01$), with DRV ($p < .001$), DRH ($p < .001$), and LN (small positive correlation, $p < .05$);
- SART-S - potential small positive correlation between averaged scenarios (avg, $p < .1$)[16], with DRV ($p < .05$), and potentially DRH ($p < .1$)[16];
- SART-Total - small negative correlation between averaged scenarios (avg, $p < .05$), with DRV (medium negative correlation, $p < .01$), DRH ($p < .05$), and LN ($p < .05$);
- NASA-TLX - medium positive correlation between averaged scenarios (avg, $p < .001$), with DRV ($p < .001$), DRH (potential small positive correlation, $p < .01$), potentially LN ($p < .1$)[16], and LD ($p < .05$);

There were no significant or even potentially trending correlations for TS.

Pearson's correlation only considers linear correlation; therefore, to verify previous findings, Spearman's correlation was also performed, which considered monotonous relationships, where, for example, the relation could be linear at first and then saturate. The results with exact values of ρ and p are shown in Tab. A.5 in the appendix. There were fewer significant correlations as compared to Pearson's correlation results: SS only showed small to medium positive correlations in SUS-L, in avg ($p < .05$), LN ($p < .05$), and LD ($p < .01$); small positive correlation in NASA-TLX, in avg ($p < .05$), potentially DRV ($p < .1$)[16], DRV ($p < .05$), and potentially LN ($p < .1$)[16]; and small negative correlation in PS, in DRV ($p < .01$), potentially DRH ($p < .1$)[16,17], and

---

[17] Marginally non-significant; exact significance was $p = .05001$.

potentially LN ($p < .1$)[16]. Contrary to Pearson's correlation, Spearman's did not show a significant correlation in SS for CHR scores, but instead, significant PS results were observed here, as opposed to Pearson's. SUS-L and NASA-TLX correlations occured in both correlations.

GP had a significant small to medium positive correlation in SART-D, in avg ($p < .01$), DRV ($p < .001$), DRH ($p < .05$), and LN ($p < .05$); potential trending small positive correlation in SART-S[16], in avg ($p < .1$), DRV ($p < .1$), and LN ($p < .1$); small to medium positive correlation for NASA-TLX, in avg ($p < .05$), DRV ($p < .01$), DRH ($p < .05$), and LN ($p < .05$); small negative correlation in PS, in potentially DRV ($p < .1$)[16], DRH ($p < .05$), and LD ($p < .05$). Compared to Pearson's Correlation of GP, Spearman's significant results did not include CHR, SUS (with sub-dimensions), SART-U, SART-D, and SART-Total. Both correlation methods included SART-D, SART-S, NASA-TLX and PS correlations.

TS had only a potential small to medium positive correlation in NASA-TLX, with avg ($p < .1$), DRH ($p < .01$), and LN ($p < .05$); and medium negative correlations in PS, with DRV ($p < .01$), DRH ($p < .01$), and LN ($p < .01$). For TS, only PS correlation was convergent with Pearson's and even showed a stronger relationship (apart from LD scenario).

In summary, there were small to medium potential trending and significant correlations of age and all of the measures for GP controller (positive correlation in CHR[18], SART-D, SART-S and NASA-TLX, and negative correlations in SUS[18], SART-U[18], SART-Total[18], and PS). It indicates that younger people tend to experience less workload using GP and have more understanding and capacity to learn this inceptor. However, it was suspected that there was an influence of a hidden factor, analysed below.

In order to unveil if there were hidden factors that influence the effect of age, univariate ANOVAs were performed where age was used as an output (dependent) variable, with Bonferroni confidence interval adjustment in pairwise comparisons and Bonferroni and Tukey HSD post-hoc multiple comparisons tests. The ANOVAs did not show significant interaction between the age and gender, handedness, TS attitude, and inceptor order. The following subsections present results from groups that showed significant effects.

---

[18] Only in Pearson's Correlation.

Gaming frequency and video game group

There was a significant main effect of gaming frequency on age, $F(4, 69) = 2.740$, $p < .05$, $\eta_p^2 = .137$. Pairwise comparisons and post-hoc tests revealed, however, that the only significant difference in age could be seen between participants who answered "yes - less than 3 hours per week" and "no/hardly ever", with mean difference $MD = -11.91$ ($p < .05$). Next analysis, using VG group and age, revealed higher significance but a smaller effect size of the main effect: $F(2, 71) = 4.958$, $p < .01$, $\eta_p^2 = .123$. Pairwise comparisons revealed that there were significant differences in age between VG groups A (frequent players) and C (non-players), $MD = -7.71$ ($p < .05$), and groups B (moderate players) and C, $MD = -10.02$ ($p < .01$). The difference between groups A and B was non-significant. Post-hoc multiple comparisons confirmed those results.

Gamepad usage

ANOVA results have shown that there was a significant main effect of gamepad usage on age, $F(3, 70) = 2.843$, $p < .05$, $\eta_p^2 = .109$, but pairwise comparisons and post-hoc tests did not reveal any significant differences between the groups. However, by observing the descriptive statistics, it can be seen that the average age for people who never used the gamepad was 36, and people who were using the gamepad at least sometimes were 29 years old, as seen in Tab. 5.9. On the other hand, the median for all groups was almost the same, which was approximately 27 years old. This explained the lack of significant differences in pairwise comparisons and post-hoc analyses.

Table 5.9: Descriptive statistics showing mean M and standard deviation SD for age as a dependent variable and GP usage as a grouping factor.

FE group	M	SD	N
yes - a lot	29.00	4.29	12
yes - sometimes	29.15	6.23	20
yes - hardly ever	28.93	6.97	14
no	36.07	14.26	28
Total	31.70	10.42	74

Mobile games usage

There was a significant main effect of MG usage on age, $F(2, 71) = 10.253$, $p < .001$, $\eta_p^2 = .224$. Here, a distinct pattern could be seen that people who never or hardly ever played MG were significantly older: from pairwise comparisons, verified by post-hoc tests, the difference between people from MG group "no/hardly ever" and "used to" was $MD = 10.03$ ($p < .01$[19]), and the difference between "no/hardly ever" and "yes" was $MD = 9.93$ ($p < .001$). There was no significant difference between "yes" and "used to". Descriptive statistics are shown in Tab. 5.10.

Table 5.10: Descriptive statistics showing mean M and standard deviation SD for age as a dependent variable and mobile games usage as a grouping factor.

FE Group	M	SD	N
yes	27.71	4.85	24
used to	27.60	4.64	20
no / hardly ever	37.63	13.38	30
Total	31.70	10.42	74

Flight experience group

From descriptive statistics in Tab. 5.11 and Fig. 5.12, it can be seen that the high-experienced pilots were older on average than low-experienced and naïve pilots. The differences were significant, with the FE group effect on age being $F(2, 71) = 8.814$, $p < .001$, $\eta_p^2 = .199$. Pairwise comparisons showed significant age differences between groups A and B: $MD = 7.44$ ($p < .05$), and A and C: $MD = 11.30$ ($p < .001$), but not between B and C. Post-hoc multiple comparisons confirmed those results.

Summary

Based on the findings in this section, it could be assumed that the age factor was a hidden factor within participants' groups. However, such analyses demonstrate that in research, every factor should be closely investigated because it may influence the final results.

---

[19] Very close to $p < .001$, as the exact value was $p = .001124$

Table 5.11: Descriptive statistics showing mean M and standard deviation SD for age as a dependent variable and flight experience (FE) group as a grouping factor.

FE group	M	SD	N
A (high experience)	38.80	13.28	20
B (low experience)	31.36	11.00	22
C (no experience)	27.50	3.81	32
Total	31.70	10.42	74

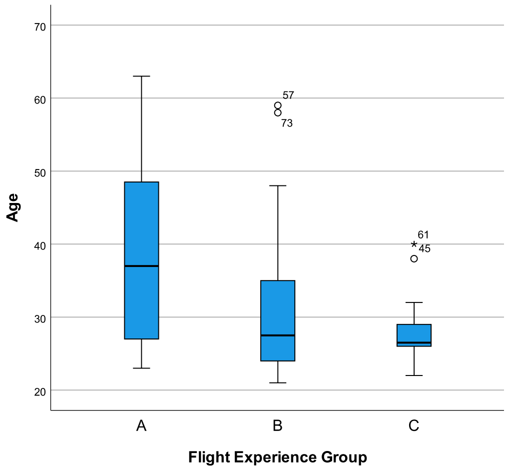


Table 5.12: Box-whisker plot showing Age distribution within flight experience (FE) groups.

5.7 Analyses of variance

This section investigates the effect of factors that were highlighted in Tab. 5.8 from Section 5.6. To inspect the degree of impact they had on the results, a factorial repeated-measures ANOVA (rANOVA) was performed for each of them. Mauchly's Test of Sphericity verified if there was a violation of covariance's homogeneity (also known as "assumption of sphericity") [124]; this assumption is met when $p > .05$. All pairwise comparisons were made with the Bonferroni adjustment, and post-hoc tests were made with equal variances assumed by Bonferroni and Tukey HSD. The effect sizes were quantified by partial eta squared (η_p^2). Partial eta squared is commonly used to quantify the proportion of total variability in the dependent variable that can be explained by a particular independent variable. It ranges from 0 to 1; a value greater than 0.01 means the effect size is small; over 0.06 is moderate; and greater than 0.14 is considered large [80]. Each subsection was concluded with a summary of the findings.

5.7.1 CHR: Cooper-Harper Rating Scale

CHR scale did not have any significant interactions with the grouping factors, so the analysis of the results was based on a 1-way rANOVA. The assumption of sphericity was not violated. The within-subject effect of inceptor type was significant: $F(2, 144) = 39.134$, $p < .001$, $\eta_p^2 = .352$. Descriptive statistics and estimates are shown in Tab. 5.13, and the score estimates are visualised on a bar graph in Fig. 5.7. Pairwise comparisons showed over 2-point CHR increase of TS over SS and GP ($p < .001$)[20]. SS and GP scores were not significantly different. Detailed results can be found in Tab. 5.14.

These results mean that participants felt a higher workload when using TS, as compared to SS and GP. The mean difference between TS and the two other inceptors was approximately 2.2 points; overall, all three inceptors in the trial were categorised as "high workload" (4-6 score in CHR) to "very high workload" (7-10 score in CHR) on average[21], but the TS was perceived as the controller most difficult to use.

Table 5.13: Descriptive statistics and estimates for CHR scores. M - mean; SD - standard deviation; N - number of samples; SE - standard error; LB - lower bound; UB - upper bound. The value of M was the same for descriptive statistics and estimates. LB and UB are in a 95% confidence interval.

	Descriptives			Estimates		
	M	SD	N	SE	LB	UB
SS	4.50	2.07	73	.242	4.02	4.99
GP	4.55	1.86	73	.218	4.12	4.99
TS	6.78	1.94	73	.228	6.33	7.24

Table 5.14: Pairwise comparisons based on estimated marginal means for CHR scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

		MD	SE	*p*	LB	UB
SS	GP	-.048	.283	1.000	-.742	.646
	TS	-2.281	.330	< .001	-3.089	-1.473
GP	SS	.048	.283	1.000	-.646	.742
	TS	-2.233	.267	< .001	-2.888	-1.578
TS	SS	2.281	.330	< .001	1.473	3.089
	GP	2.233	.267	< .001	1.578	2.888

---

[20] There was a significant difference between TS and SS ($MD = 2.28$, $SE = 0.33$, $p < .001$), and TS and GP ($MD = 2.23$, $SE = 0.27$, $p < .001$).

[21] With SD taken into consideration, some scores of SS and GP have also been categorised as "low workload" category.

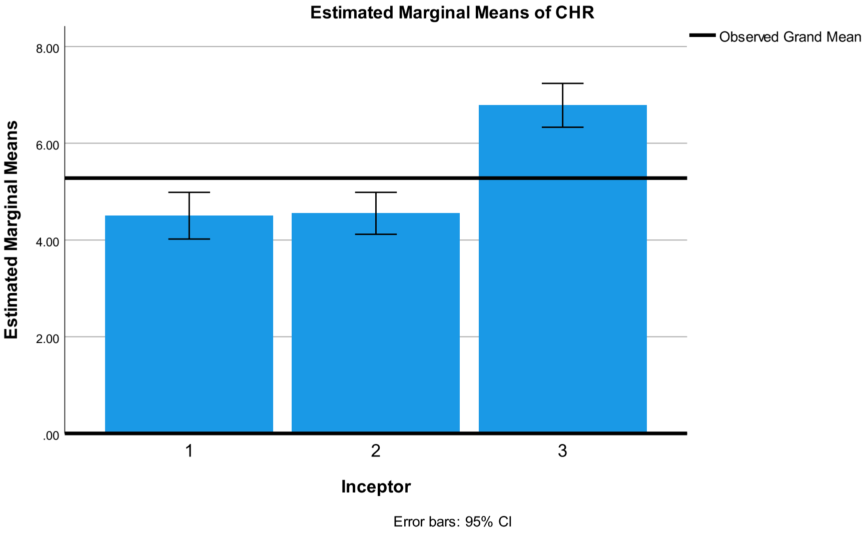


Figure 5.7: Estimated marginal means for CHR. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.2 SUS-U: System usability scale - usability

Previous analysis showed that two grouping factors had a significant effect on SUS-U results among participants: TS attitude, which was the participant's attitude towards introducing TS technology in flight decks, and FE group. A trend of the VG group effect was also reported. To further investigate this interaction, three separate factorial rANOVAs were carried out. SUS-U scores across the three inceptors were a repeated-measures dependent factor, and each group type was an independent factor.

Touchscreen attitude as an independent factor

The TS attitude factor grouped participants by their answer to the question "What is your view on touchscreen technology being introduced in aircraft cockpits? Use the scale from 1 (I do not like the idea) to 5 (I like the idea)" in the demographic questionnaire. Descriptive statistics are presented in Tab. 5.15, and Fig. 5.8. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. The main effect of the inceptor was significant: $F(2, 134) = 62.187$, $p < .001$, $\eta_p^2 = .481$, but there was no interaction between the inceptor and TS attitude. Between-subject effect of TS attitude was also not significant.

Pairwise comparisons and post-hoc trials only confirmed that TS attitude did not influence the experienced usability of the controller; there were no significant interactions, overall and within each inceptor. The main effect of the inceptor was/is? analysed in a later section. Still, it is worth noting that patterns within each TS attitude were the same among each group, which was convergent with the overall inceptor comparisons.

Data from the analysis can be found in Tab. A.6, A.7, A.8, A.9, and A.10, with bar plot of SUS-U mean scores for each inceptor and TS attitude group shown in Fig. 5.8.

In conclusion, even though the mean results showed somewhat better scores among people who responded with ”5 (I like the idea)” to the TS attitude question, there were no visible patterns in other scores. Therefore, this higher score was assumed to be coincidental, especially since the difference was not significant. Based on the results in this section, it was decided to exclude the TS attitude grouping factor in later SUS-U analyses.

Table 5.15: Descriptive statistics for SUS-U scores with TS attitude as a grouping factor. The scale was from 1 (negative attitude) to 5 (positive attitude). M - mean; SD - standard deviation; N - number of samples.

TS attitude	SS		GP		TS		N
	M	SD	M	SD	M	SD	
1	54.93	14.16	50.49	21.43	26.53	22.85	9
2	55.69	19.50	52.99	13.40	29.93	23.45	9
3	54.93	11.58	52.47	15.53	25.45	14.24	18
4	58.92	9.44	51.81	13.49	30.52	12.80	18
5	60.35	12.19	54.79	16.37	38.92	15.35	18
Total	57.38	12.65	52.70	15.48	30.78	17.07	72

Flight experience group as an independent factor

This factor included three groups, based on participant’s FE: A (high-experienced pilots), B (low-experienced pilots), and C (naïve pilots). Descriptive statistics are presented in Tab. 5.16 and visualised in Fig. 5.9. Mauchly’s test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 85.326$, $p < .001$, $\eta_p^2 = .549$), FE group ($F(2, 70) = 5.822$,

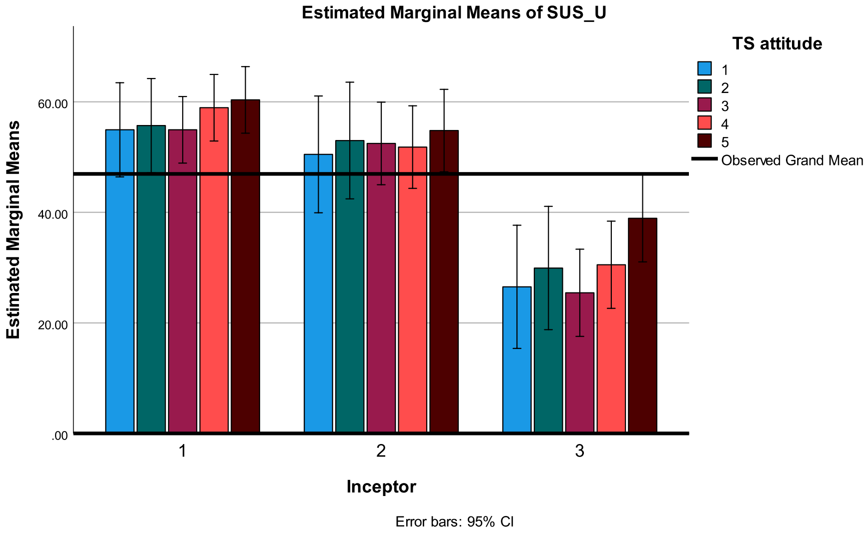


Figure 5.8: Estimated marginal means for SUS-U within participant's TS attitude. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

$p < .01$, $\eta_p^2 = .143$), and the interaction between inceptor and group ($F(4, 140) = 4.300$, $p < .01$, $\eta_p^2 = .109$).

Pairwise comparisons of FE groups showed significant SUS-U score differences between groups A and B ($p < .01$) and A and C ($p < .05$) - participants from group A gave lower ratings by 7-9 points on average. There was no significant difference between B and C groups' scores, which was confirmed with post-hoc analysis. The overall scores between all three inceptors were significantly different ($p < .05$ and less), with estimated means of $M = 57.754$ for SS, $M = 51.728$ for GP, and $M = 29.102$ for TS. In the case of SS, the only significant interaction could be seen in group B, with the SUS-U score being 11 points higher than in group C ($p < .01$), while for GP, in group A, ratings were 10 points lower than in group C ($p < .05$). More interactions could be found in TS: scores from group A were 12 points lower than from group B ($p < .1$[22]), and 16 points lower than group C ($p < .01$).

Pairwise comparisons between inceptors within each FE groups A and B revealed significant differences between every inceptor ($p < .05$ and less). For group C, the only non-significant difference was between SS and GP. In groups A and B, scores for SS were approximately 10-11 points higher than GP and 32-37 points higher than TS. GP had approximately a 21-to-25-point advantage over TS. Within group C, GP appeared to be the highest-rated inceptor; however, the mean difference of approximately 3.5

---

[22] Very close to $p < .05$; actual value was $p = .055$.

points over SS was non-significant. The scores for TS in this group were lower than SS and GP by 16 and 20 points, respectively.

Data from the analysis can be found in Tab. A.11, A.12, A.13, A.14, and A.15, with bar plots of SUS-U mean scores for each inceptor and FE group shown in Fig. 5.9.

The SUS-U scores could also be interpreted using the scale provided in Section 4.7.3. Mean scores for SS and GP across all FE groups, ranging between $M = 45.39$ (GP, group A) and $M = 63.92$ (SS, group B), indicated that the usability of those controllers has been recognised as "good" on the adjective ratings scale[23]. SUS-U scores for TS depended on a FE group - the usability for high-experienced pilots (group A) could be ranked as "worst" on average, while for FE groups B and C - "poor"[24].

Using the acceptance scale definitions, most participants rated SS at an "acceptable" usability level [25]. GP was rated mostly as "marginal low" by group A (even though the task performance with this controller was often better than with SS, as seen in PS score analyses and results in Section 5.7.15), "marginal high" by FE group B, and "acceptable" by group C. [26] TS's usability was deemed as "not acceptable" in most cases across all three FE groups[27].

In conclusion, FE played a significant role in the scoring of the inceptor's usability. High and low-experienced pilots have rated SS significantly higher than GP. This was due to the pilot training and familiarity with SS. However, based on the scale from Section 4.7.3, the usability of both controllers could be interpreted as "good". In contrast, only non-pilots gave slightly higher scores to GP over SS, albeit with a non-significant difference. This implies that SS and GP were similarly usable for them. The average SUS-U rating of TS from high-experienced pilots was categorised as "worst" usability, while the score from the other two groups could be described as "poor".

---

[23] Taking SD into consideration, some higher scores of SS and GP reached "best" and "excellent" usability, while the lower scores reached "OK / fair" description, and in some GP cases also "poor".

[24] Again, with SD and for group A, TS usability was ranked as "poor" and "OK / fair", and for groups B and C, it was ranked up to "good"

[25] FE group A results included levels down to "marginal low", group B results were down to "marginal high", and some group C results even included "not acceptable" usability rating.

[26] Ratings for groups A and B ranged from "not acceptable" up to "acceptable"; for group C, ratings started at a "marginal low" level and did not reach more than the reported "acceptable".

[27] Although it was reaching up to "marginal high" usability levels in groups B and C.

Table 5.16: Descriptive statistics for SUS-U scores with FE group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Flight Exp.	SS		GP		TS		N
Group	M	SD	M	SD	M	SD	
A	56.64	11.88	45.39	16.17	19.67	12.42	19
B	63.92	10.52	53.64	16.26	31.93	18.11	22
C	52.70	13.00	56.15	13.18	35.70	16.73	32
Total	57.11	12.77	52.59	15.39	30.39	17.27	73

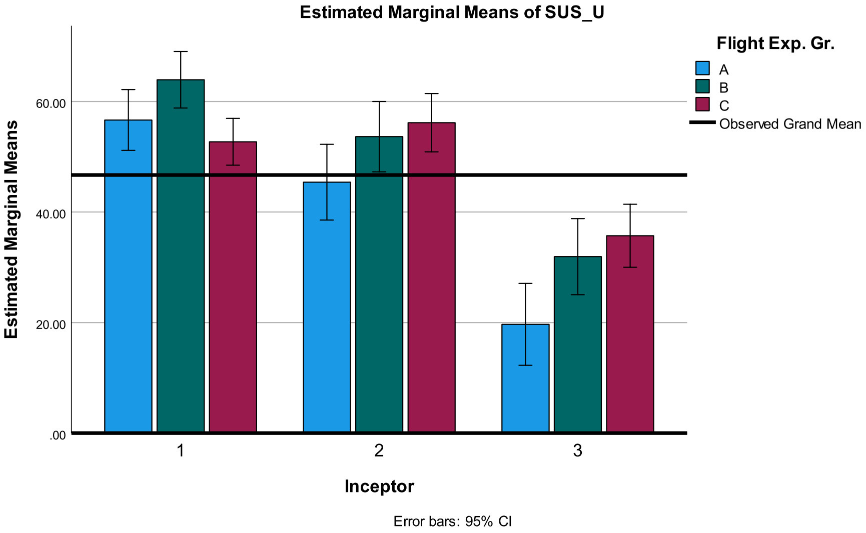


Figure 5.9: Estimated marginal means for SUS-U across FE groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Video game group as an independent factor

This factor included three groups: A (high-experienced gamers), B (low-experienced gamers), and C (non-gamers). Descriptive statistics are presented in Tab. 5.17, and Fig. 5.10. The assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2,140) = 60.867$, $p < .001$, $\eta_p^2 = .465$), but the VG group and its interaction on the inceptor were not significant.

There were no significant results in pairwise comparisons and post-hoc tests of VG groups; however, VG groups A and B rated the usability of inceptors marginally higher than group C (by approximately 5-6 points, non-significantly). There were significant differences in pairwise comparisons based on estimated marginal means, similar to those in the section with FE group as an independent factor (page 119). The estimated means were $M = 56.689$ for SS, $M = 50.898$ for GP and $M = 30.324$ for TS.

There was only one significant pattern in Inceptor * VG group interaction analysis: for GP, VG group C placed the GP significantly lower (by 12-15 points, $p < .05$ and lower).

Within VG groups A and B, there are significant differences between SUS-U scoring of SS and TS: SS scores had 27 more points in groups A and B ($p < .001$), and between GP and TS: GP scores were 23-26 higher in groups A and B, respectively ($p < .001$). VG group C had a significant difference in ratings of SS and GP (12 more points for SS, $p < .05$) and SS and TS (24-point advantage of SS, $p < .001$), but the fact that GP had 11 points more than TS was not significant.

Data from the analysis can be found in Tab. A.16, A.17, A.18, A.19, and A.20, with bar plots of SUS-U mean scores for each inceptor and VG group shown in Fig. 5.10.

Based on adjective and acceptance scales, SS was rated at "good" and "acceptable" (groups A and B) / "marginal (high)" (group C) usability levels, respectively[28] GP scored slightly lower: "good" on the first scale within all three groups; and "marginal (high)" in group A, "acceptable" in group B, and "marginal (low)" in group C on the second scale[29]. TS was deemed "OK / fair" by group A (by just 0.15 points over the lower category), and "poor" by groups B and C (by 0.01-0.36 points below the higher category). On the second scale, the results fell into the "not acceptable" category across all three groups[30].

Overall, the usability scores tended to be similar across the VG groups, but gamers (with any experience) tended to give higher usability scores for SS and GP, as compared to non-gamers. This could especially be seen in the case of GP, where the non-gamers scores were approximately 12-15 points lower on a SUS-U scale than those who play VGs. This showed that people who were familiar with a controller (in this case: GP

---

[28] Within SD range from the mean score, SS also reached "best" (groups A and B) and "excellent" (group C) categories; and ranged from "marginal (low)" to "acceptable" category (all three groups).

[29] In group A, the scoring ranged from "OK / fair" to "excellent", and "marginal (low)" to "acceptable"; group B included scales from "OK / fair" to "best", and also "marginal (low)" to "acceptable"; for group C, the first scale was in the range of "poor" to "excellent", while the acceptance scale ranged across all steps.

[30] With the range from "worst" to "good" across all groups, and up to "marginal (low)" for groups A and B / "marginal (high)" for group C.

for gamers) were more comfortable using it. TS controller was almost evenly scored by participants as a "not acceptable" controller in its current form, which is understandable, given its first iteration and hardware limitations. It also indicated that the VG experience did not influence the reception of an unfamiliar controller.

Table 5.17: Descriptive statistics for SUS-U scores with VG group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

VG Group	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	57.55	13.88	53.75	12.45	30.54	18.49	35
B	57.92	12.25	57.29	15.71	30.39	14.35	24
C	54.60	11.21	41.65	17.24	30.04	19.83	14
Total	57.11	12.77	52.59	15.39	30.39	17.27	73

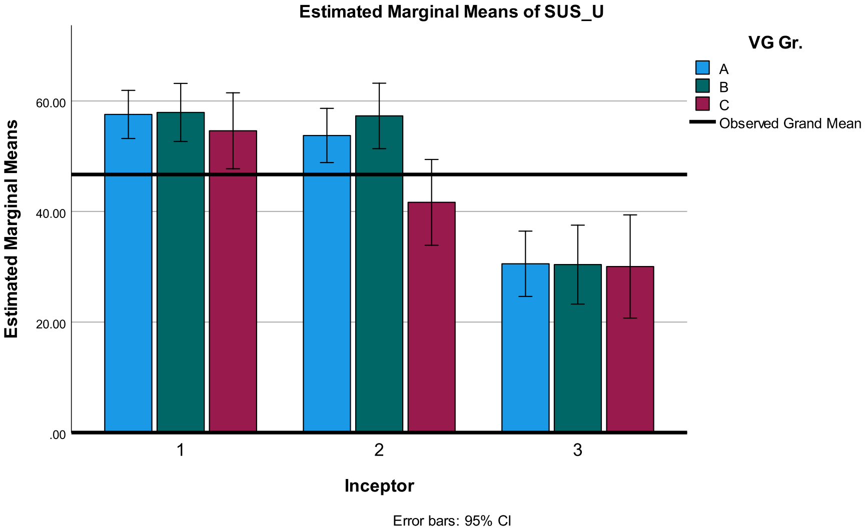


Figure 5.10: Estimated marginal means for SUS-U within participant's VG group. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Summary

Analyses in this section showed that the most influential factor in the inceptor usability scores (SUS-U) was the flight experience (FE). Video game (VG) experience was meaningful in GP ratings, but it was found that it did not influence the reception of TS, which was an unfamiliar controller to every participant. Moreover, the higher usability scoring of the SS among the non-gamers could be explained by the fact that the distribution of participants with high flight experience in this group was higher, which can be seen in Fig. 5.11.

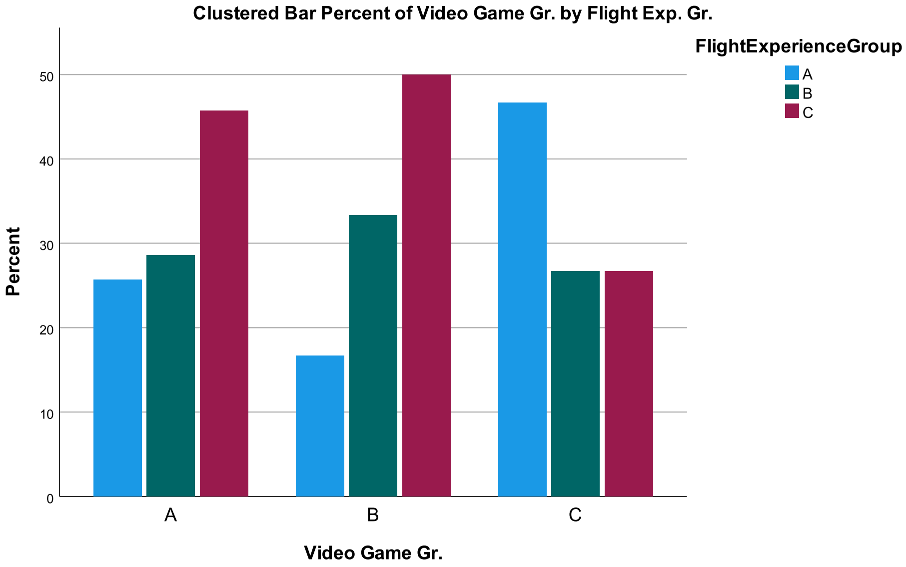


Figure 5.11: Distribution of FE groups across each VG group. Percentage within each group cluster sums up to 100%.

Overall scores can be seen in Fig. 5.12. Using the adjective and acceptance categorisation system, developed in Section 4.7.3, the usability of SS and GP was mostly rated as "good" (adjective scale), and "marginal (high)" / "acceptable" (acceptance scale), while TS's usability was rather rated between "poor" and "OK / fair" on one scale, but "not acceptable" on the other scale.

5.7.3 SUS-L: System usability scale - learnability

The learnability of the controller was measured with the SUS-L scale. Earlier analyses indicated a small effect of gender, mobile games, and VG experience on learnability. This section investigates the effect of those grouping factors using three separate 2-way mixed-design ANOVAs. SUS-L scores across the three inceptors were a repeated-measures dependent factor, and each group type was an independent factor.

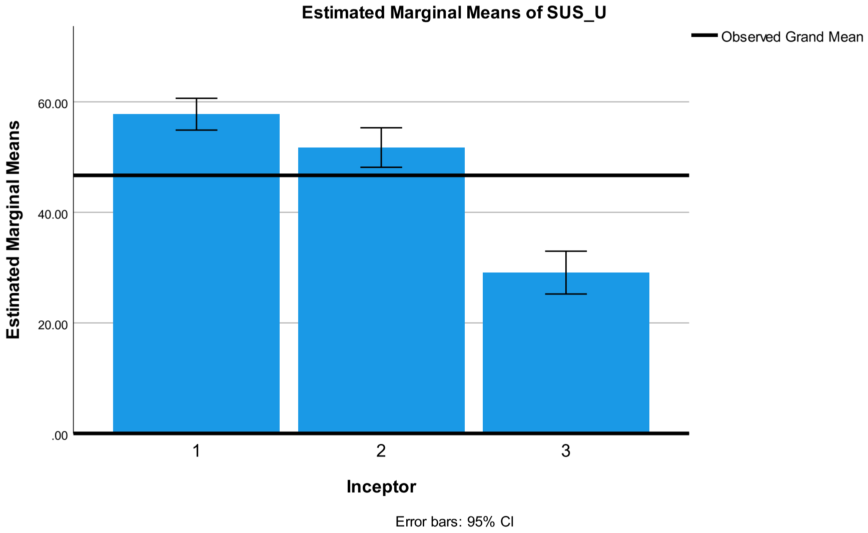


Figure 5.12: Estimated marginal means for SUS-U. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Gender as an independent factor

This section investigates the effect of gender on SUS-L scores. Descriptive statistics are presented in Tab. 5.18, and Fig. 5.14. The assumption of sphericity was not violated. There were no significant effects of any of the factors or interactions.

Pairwise comparisons for the main effects of gender and inceptors did not show significant patterns. In the interaction analyses, there was a similar pattern for both male and female participants - GP SUS-L scores were significantly higher as compared to TS by approximately 4 points for females ($p < .01$) and 5 points for males ($p < .001$). Additionally, SS results were 4 points higher than TS among male participants ($p < .001$). Although, given the uneven amount of samples within each group, this could be coincidental.

Data from the analysis can be found in Tab. A.21, A.22, A.23, and A.24.

Expectedly, this factor did not have a significant effect on the SUS-L scores. The possible reason why it was found to be a potentially significant factor is that the distribution of genders was similar within VG groups, which can be seen in Fig. 5.13. Based on the results in this section, it was decided to exclude the gender factor in later SUS-L analyses.

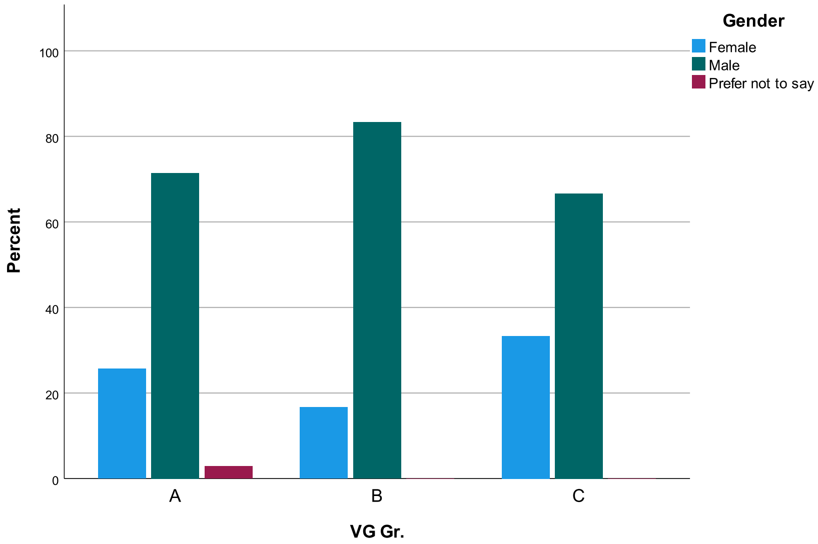


Figure 5.13: Distribution of gender across each VG group. Percentage within each group cluster sums up to 100%.

Table 5.18: Descriptive statistics for SUS-L scores with gender as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Gender	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Female	11.98	4.61	13.92	6.02	10.03	6.34	18
Male	14.73	4.50	15.32	4.66	10.51	5.07	54
Prefer not to say	20.00		20.00		20.00		1
Total	14.13	4.67	15.04	5.03	10.52	5.45	73

Mobile games usage as an independent factor

Grouping by mobile games (MG) usage was done by three categories (which answered the question ”do you play mobile games on your smartphone/tablet?”): ”no / hardly ever”, ”used to”, and ”yes”. The main interest of analysis in this section was to check if mobile gaming experience influenced TS results. Descriptive statistics are presented in Tab. 5.19, and Fig. 5.15. The assumption of sphericity was not violated. The rANOVA results showed that there was a significant effect of inceptor: $F(2, 140) = 35.091$, $p < .001$, $\eta_p^2 = .334$, and of the inceptor * MG interaction:

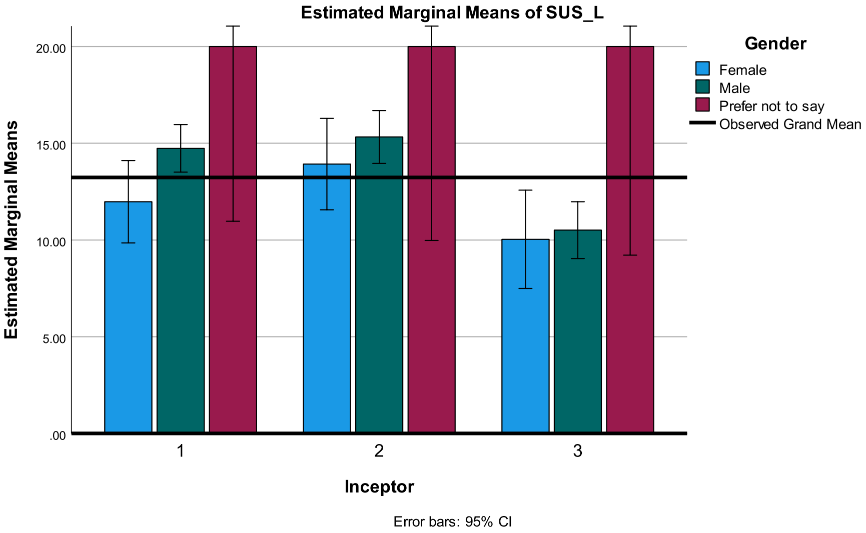


Figure 5.14: Estimated marginal means for SUS-L within participants' gender. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

$F(4, 140) = 6.976$, $p < .001$, $\eta_p^2 = .166$. Between-subject effect of MG was not significant.

Pairwise comparisons of scores grouped by MG usage did not show any significant differences. There were significant differences between the inceptors: SS and GP had significantly higher SUS-L scores than TS (by approximately 3-4 points on average, $p < .001$). There was no significant difference between SS and GP; however, they showed a trend when the tolerance was slightly elevated: SS results were lower than GP, but only by 1.2 points ($p < .1$[31]). The interaction analysis showed two significant relationships between MG usage groups, albeit both for SS, which was deemed coincidental and not relevant in this case.

Results from the analysis in this section can be found in Tab. A.25, A.26, A.27, and A.28.

Results from this section show that different mobile gaming usage among participants did not influence the perceived learnability of any inceptor, so it was decided to exclude this factor in later SUS-L discussions.

---

[31] Actual value was $p = .054$, so only marginally higher than required $\alpha = .05$.

Table 5.19: Descriptive statistics for SUS-L scores with MG usage as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

MG usage	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no / hardly ever	16.51	3.16	13.79	5.78	11.79	6.18	29
used to	11.78	6.00	14.63	5.15	9.16	4.74	20
yes	13.20	3.69	16.90	3.30	10.13	4.92	24
Total	14.13	4.67	15.04	5.03	10.52	5.45	73

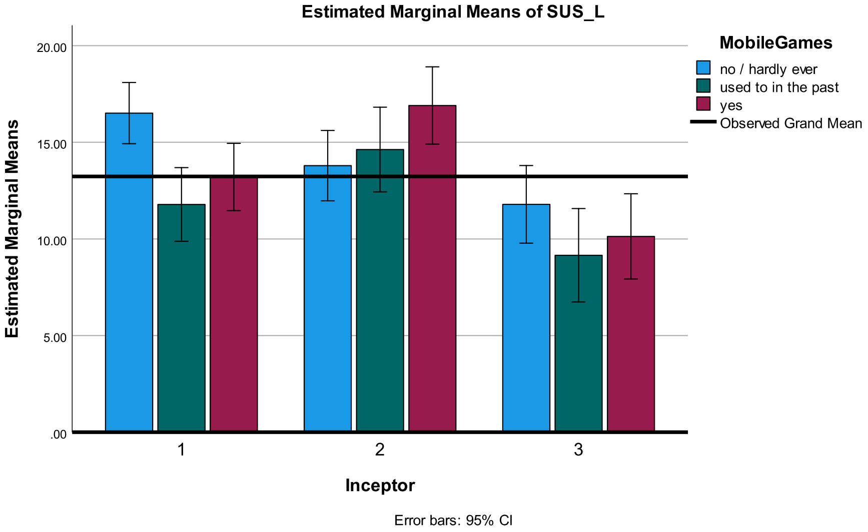


Figure 5.15: Estimated marginal means for SUS-L within participant's MG usage. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Video game group as an independent factor

This section investigates the effect of VG group on SUS-L scores. Descriptive statistics are presented in Tab. 5.20, and Fig. 5.16. The assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2,140) = 25.957$, $p < .001$, $\eta_p^2 = .271$), but the VG group and its interaction on the inceptor were not significant.

There were no significant interactions between VG groups in pairwise and post-hoc comparisons. Inceptor estimates showed a significant 4-point disadvantage of TS compared to SS and GP ($p < .001$).

Interaction analysis for the inceptor * VG group did not reveal significant differences in SS and TS scores. There were two interactions among GP results: group C ranked GP's learnability 4.6 points lower than groups A ($p < .01$) and C ($p < .05$). Furthermore, there were some interactions within each group: in groups A and B, SS and GP scored significantly higher than TS on the SUS-L scale (by approximately 3-5 points, $p < .05$ and less); in group C, the only significant difference was between SS and TS (by approximately 5 points, $p < .01$).

Data from the analysis can be found in Tab. A.29, A.30, A.31, A.32, and A.33, with bar plots of SUS-L mean scores for each inceptor and VG group shown in Fig. 5.16.

From the adjective SUS-L scales definition, the learnability of SS could be reported as "excellent" (indicated by VG group A) and "good" (groups B and C)[32]. Learnability score of GP could be interpreted as "excellent" (groups A and B) and "good" (group C)[33]. In the case of TS, participants ranked its learnability as "good" (group A) and "OK / fair" (groups B and C) on average[34].

In conclusion, there seems to be a small link between VG experience and learnability scores. Non-gamers tended to rate the GP and TS lower than gamers, who gave very similar learnability ratings for all three inceptors. Expectedly, GP was rated best by gamers, making it a controller with the biggest mean difference when compared to non-gamers ratings. The analysis also showed that there was a (non-significant) tendency for gamers to learn a novel controller (TS) quicker than non-gamers.

---

[32] Results from all VG groups ranged from "OK / fair" to "best".

[33] Groups A and B scales ranged from "good" to "best", however, learnability reports from group C ranged across the whole range ("worst" to "best").

[34] Ratings of group A included ranks from "poor" to "best"; group B - from "poor" to "good"; and group C - from "worst" to "excellent".

Table 5.20: Descriptive statistics for SUS-L scores with VG group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

VG Gr.	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	14.68	4.92	15.95	4.19	11.52	5.69	35
B	13.44	4.77	15.91	4.08	10.23	4.27	24
C	13.93	3.96	11.29	6.78	8.53	6.38	14
Total	14.13	4.67	15.04	5.03	10.52	5.45	73

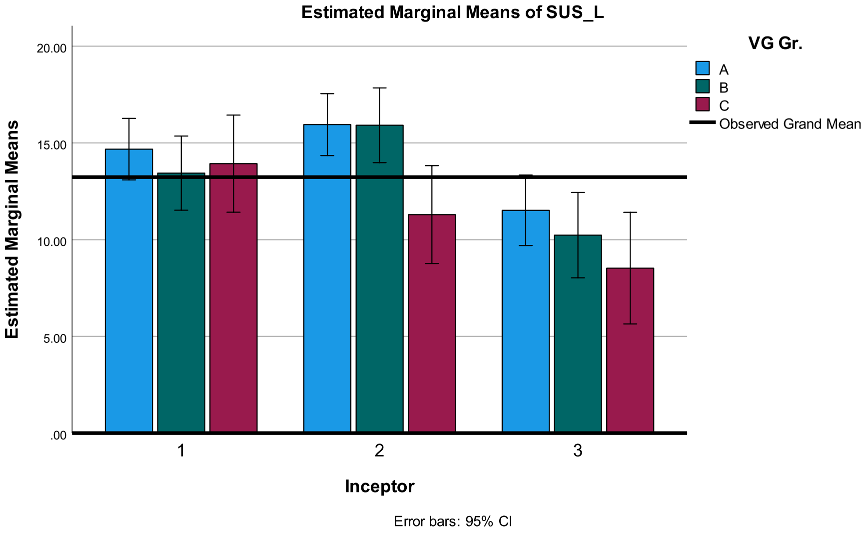


Figure 5.16: Estimated marginal means for SUS-L with VG group. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Summary

Investigation of independent factors in this section demonstrated that gender and mobile games usage did not have a significant influence on the learnability ratings (SUS-L). The fact that the gender factor was marked as influential in Section 5.6 might indicate that it was a ”hidden factor” in other groups. For example, the majority of experienced pilots were males, which might have caused the ANOVA to detect gender as a significant factor. There is some potential in exploring VG experience as a factor in the learnability of new controls in future pilot training research; however, in this research, the link was not strong. The reason for this might be that other factors, such as flight experience, largely influenced the analyses.

Additional 1-way rANOVA (with no independent factors) showed that there was a significant learnability difference between the TS and two other inceptors, but no significant difference was reported between SS and GP (Fig. 5.17). Nonetheless, placed on a scale from Section 4.7.3, SS and TS could be qualified as inceptors with "good" learnability, while GP would fall into the "excellent" learnability category.

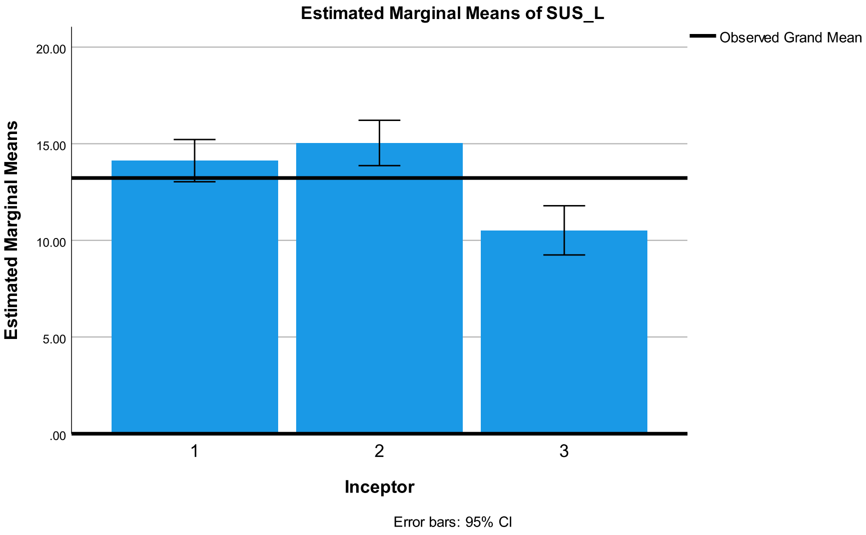


Figure 5.17: Estimated marginal means for SUS-L. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.4 SUS-Total: System usability scale - total score

Analysis in Section 5.6 revealed that gender, TS attitude, FE and VG experience might have had a slight influence on the SUS-Total results, which is a combination of usability (SUS-U) and learnability (SUS-L) scores. Four separate 2-way rANOVAs were carried out to investigate the effect of each of those groups on the SUS-Total score.

Gender as an independent factor

This section investigates the effect of gender on SUS-Total scores. Descriptive statistics are presented in Tab. 5.21, and Fig. 5.19. The assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 10.481$, $p < .001$, $\eta_p^2 = .130$), but the gender and its interaction with the inceptor were not significant.

Similar to the analyses of gender factor effect on SUS-L score, pairwise comparisons did not show significant differences between gender. Inceptor analysis revealed that SS and GP were rated better than TS by 30-33 points ($p < .001$). SS and GP scores were almost identical, so the 2-point difference between them was non-significant.

Pairwise comparisons of inceptor * gender interactions showed only one significant difference in SS: male participants rated it 10 points higher than females ($p < .05$); however, this situation could be explained similarly to those of the SUS-L scores - there were more males among gamers and experienced pilots, which can be seen in Fig. 5.13 and 5.18, respectively. Gender * inceptor interactions confirmed that SS and GP results were significantly higher than those of TS.

Data from the analysis can be found in Tab. A.34, A.35, A.36, and A.37.

In conclusion, the gender factor did not have a significant effect on the SUS-Total scores. Therefore, it was decided to exclude the gender factor in later SUS-Total analyses.

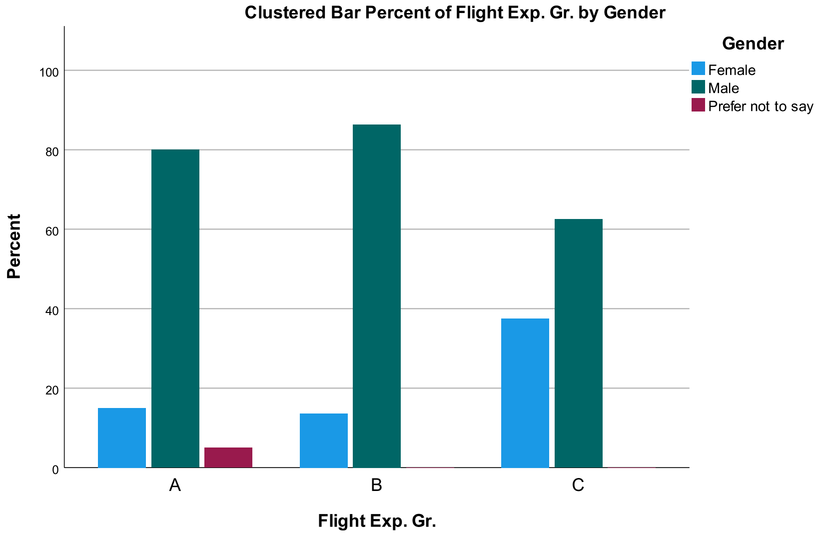


Figure 5.18: Distribution of gender across each FE group. Percentage within each group cluster sums up to 100%.

Table 5.21: Descriptive statistics for SUS-Total scores with gender as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Gender	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Female	63.30	16.93	65.31	21.77	41.11	25.13	18
Male	73.41	14.51	67.95	17.62	40.66	18.75	54
Prefer not to say	96.25		92.50		51.25		1
Total	71.23	15.83	67.64	18.72	40.92	20.23	73

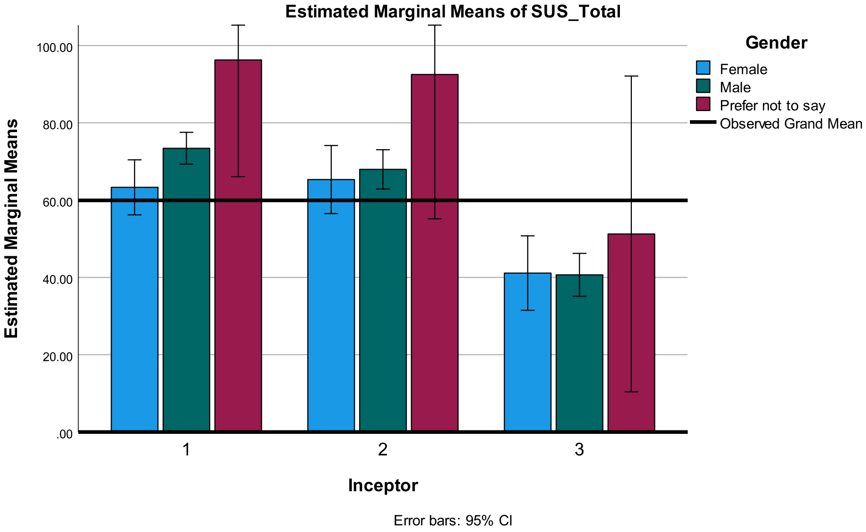


Figure 5.19: Estimated marginal means for SUS-Total within participants' gender. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Touchscreen attitude as an independent factor

This section investigates the effect of TS attitude on SUS-Total scores. Descriptive statistics are presented in Tab. 5.22, and Fig. 5.20. The assumption of sphericity was not violated. Similarly to the analysis of this factor in SUS-U results (Section 5.7.2, page 118), the main effect of inceptor was significant: $F(2, 134) = 59.692$, $p < .001$, $\eta_p^2 = .471$, but the TS attitude main effect and inceptor * TS attitude interaction were non-significant.

Pairwise comparisons showed no specific interactions between different TS attitudes, which was confirmed with post-hoc tests. Inceptor interactions were similar to those in the gender factor analysis (page 132). The pattern of SS and GP scores being significantly higher than TS scores appeared among all TS attitudes.

Data from the analysis can be found in Tab. A.38, A.39, A.40, A.41, and A.42, with bar plots of SUS-Total mean scores for each inceptor and TS attitude shown in Fig. 5.20.

In conclusion, it was decided to exclude the TS attitude factor in later SUS-Total discussions because the analyses did not reveal its significant interaction with the results, especially within TS inceptor.

Table 5.22: Descriptive statistics for SUS-Total scores with TS attitude as a grouping factor. The scale was from 1 (negative attitude) to 5 (positive attitude). M - mean; SD - standard deviation; N - number of samples.

TS att.	SS		GP		TS		N
	M	SD	M	SD	M	SD	
1	67.71	16.38	63.61	26.66	36.25	27.87	9
2	71.46	23.58	69.31	14.98	44.79	26.54	9
3	67.47	16.33	67.64	18.47	35.28	17.21	18
4	73.89	11.47	67.43	15.75	40.42	15.67	18
5	74.90	14.72	69.58	20.87	49.03	18.01	18
Total	71.46	15.82	67.78	18.82	41.31	20.09	72

Flight experience group as an independent factor

This section investigates the effect of FE group on SUS-Total scores. Descriptive statistics are presented in Tab. 5.23 and visualised in Fig. 5.21. The assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 83.541$, $p < .001$, $\eta_p^2 = .544$) and interaction between the inceptor and FE group ($F(4, 140) = 4.182$, $p < .01$, $\eta_p^2 = .107$), but not of the group alone (there was a trend when the tolerance was elevated to 10% - $F(2, 70) = 2.809$, $p < .1$, $\eta_p^2 = .074$).

Pairwise comparisons and post-hoc tests did not show significant differences between the FE groups. Nevertheless, some patterns between the inceptors and group * inceptor interactions were observed. Similar to previous rANOVAs from this section,

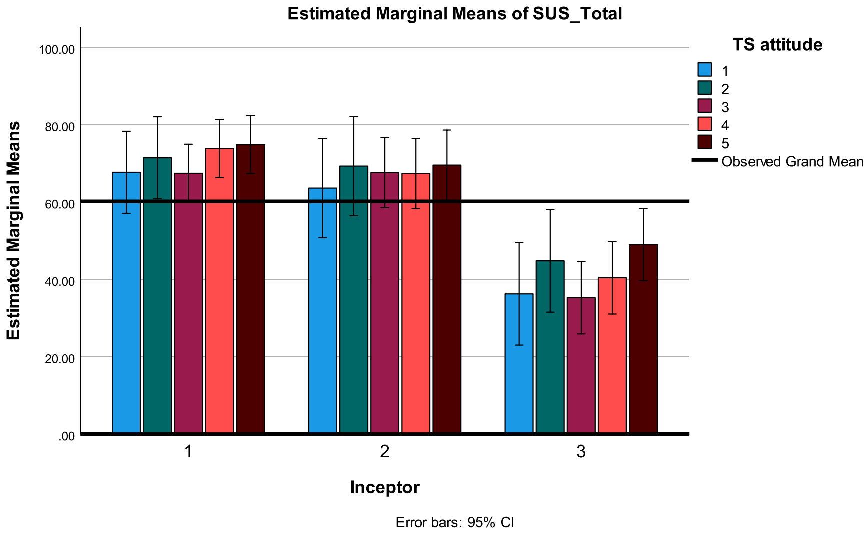


Figure 5.20: Estimated marginal means for SUS-Total within participant's TS attitude. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

SS and GP results did not differ from each other, but both had significantly higher SUS-Total scores than TS.

Further analyses have revealed that there was a significant difference between FE groups B and C in SS rating (group B rated SS approximately 13 points higher than group C, $p < .01$) and between groups A and C in TS rating (group A rated TS 15 points lower than group C, $p < .05$). The FE group * inceptor interaction showed that, among all FE groups, SS and GP were rated significantly higher than TS. GP to TS ratings had diminutive mean differences, but SS to TS score varied significantly ($p < .001$). In group A, the average SS and TS difference was the highest, reaching 42.5 points; in group B, it was equal to 35.5 points; and in group C, the difference was the lowest, approximating 19.5 points. Groups B and C did not have significant interaction between SS and GP, but group A did: SS ratings were higher by 13 points ($p < .05$).

Data from the analysis can be found in Tab. A.44, A.45, A.46, and A.47, with bar plots of SUS-Total mean scores for each inceptor and FE group shown in Fig. 5.21.

From the adjective and acceptance SUS-Total scales definitions (defined in Section 4.7.3), the system usability of SS could be reported as "excellent" / "acceptable" (indicated by FE groups A and B) and "good" / "marginal (high)" (group C)[35].

---

[35] SS results ranged from "good" / "marginal (low)" for FE group A, "good" / "marginal (high)" for group B, and "OK / fair" / "not acceptable" for group C. Top margins were "best" / "acceptable"

The SUS-Total score of GP could be interpreted as "good" (across all FE groups) / "marginal (low)" for group A, "marginal (high)" (almost "acceptable") for group B, and "acceptable" (group C)[36]. In the case of TS, participants ranked its learnability as "poor" (group A) and "OK / fair" (groups B and C) on average. All groups' SUS-Total scores placed TS as "not acceptable"[37].

Contrary to SUS-U results, flight experience appeared to have only a medium effect on the SUS-Total score. The reason for this might be that inceptors had similar SUS-L scores among all participants, and even though SUS-L contributes only to 20% of the SUS-Total score, it significantly influenced this measure.

Table 5.23: Descriptive statistics for SUS-Total scores with FE group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

FE Gr.	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	73.19	13.41	59.87	20.84	30.66	16.06	19
B	78.15	14.65	69.23	19.32	42.59	21.61	22
C	65.31	16.11	71.15	16.05	45.86	19.83	32
Total	71.23	15.83	67.64	18.72	40.92	20.23	73

Video game group as an independent factor

This section investigates the effect of VG group on SUS-Total scores. Descriptive statistics are presented in Tab. 5.24, and Fig. 5.22. The assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 61.602$, $p < .001$, $\eta_p^2 = .468$); the VG group was showing a trend when the tolerance was elevated to 10% ($F(2, 70) = 2.554$, $p < .1$, $\eta_p^2 = .468$). The inceptor * VG group interaction was not significant.

---

for groups B and C, and "excellent" / "acceptable" for group C.

[36] GP scales of groups A and B ranged from "OK / fair" / "not acceptable" (almost "marginal (low)" for group B) to "excellent" (group A) and "best" (group B). Group C ranged from "good" / marginal (low)" to "best". All groups were reaching the "acceptable" acceptance level.

[37] TS ratings of group A included ranks from "poor" to "OK / fair"; group B and C - from "worst" to "good". Acceptance within group A was only "not acceptable", but groups B and C were reaching the "marginal (high)" threshold.

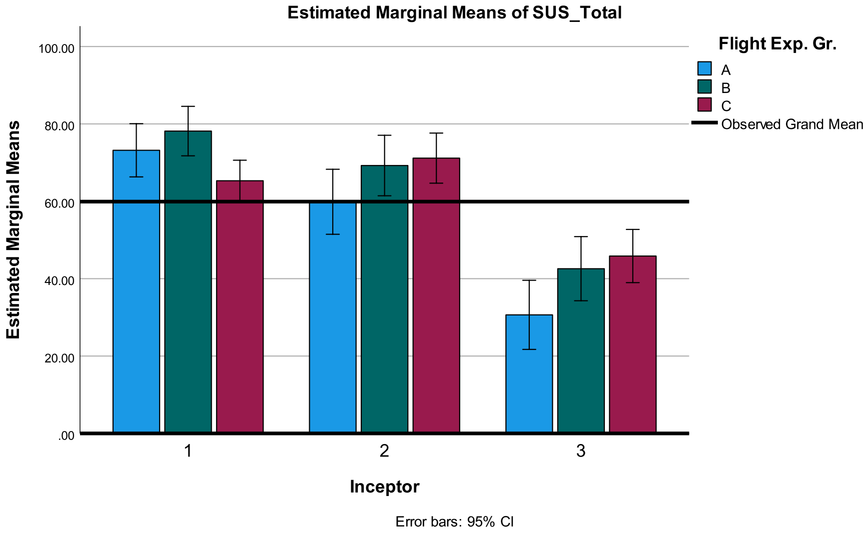


Figure 5.21: Estimated marginal means for SUS-Total across FE groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Pairwise comparisons of VG group alone did not show significant differences, although the SUS-Total estimated means within each group were $M = 61.327$ for group A, $M = 61.727$ for group B, and $M = 53.348$ for group C. Post-hoc tests were also non-significant. There were significant differences between inceptors - again, only between TS and the other two. SS and GP results had a mean difference score of 24-30 points over TS ($p < .001$).

Pairwise interactions of inceptor * VG group were only significant in the case of GP: group C scored this inceptor significantly lower than groups A (by 17 points, $p < .05$) and B (20 points of mean difference, $p < .01$). SS and TS did not show any significant differences between the VG groups. VG group * inceptor comparisons revealed more significant interactions: in VG groups A and B, there was a ”usual” high difference between SS/GP and TS; in group C, however, the interaction between GP and TS was not significant (there was a trend when the tolerance was elevated to 10%; in this case, GP had 14 points more on average than TS, $p < .1$), and there was a significant interaction effect between SS and the other two inceptors: it had a 15.5-point advantage over GP ($p < .05$), and a 30-point advantage over TS ($p < .001$).

Data from the analysis can be found in Tab. A.49, A.50, A.51, and A.52, with bar plots of SUS-Total mean scores for each inceptor and VG group shown in Fig. 5.22.

In conclusion, the only significant influence of VG experience on SUS-Total score was found to be within GP results, where gamers rated this controller significantly higher than non-gamers. This is logical, as gamers, even if they have never used the gamepad themselves, are aware of it and understand its principles. No other relevant interactions were found, so it could be assumed that the VG group did not have considerable influence on the results and could be excluded from further analysis.

Table 5.24: Descriptive statistics for SUS-Total scores with VG group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

VG Gr.	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	72.23	17.20	69.70	14.32	42.05	22.72	35
B	71.35	15.16	73.20	18.51	40.63	15.50	24
C	68.53	14.03	52.95	22.31	38.57	21.97	14
Total	71.23	15.83	67.64	18.72	40.92	20.23	73

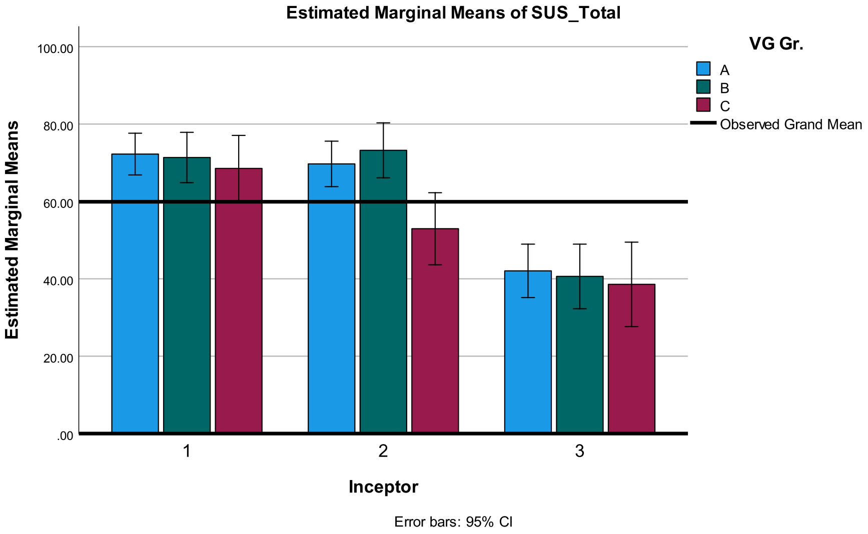


Figure 5.22: Estimated marginal means for SUS-Total across VG groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Summary

Through the rANOVA tests, it was revealed that none of the potentially significant grouping factors (gender, TS attitude, FE and VG experience) played a meaningful role in the SUS-Total scoring. Therefore, a separate 1-way rANOVA was performed to investigate the SUS-Total score differences between inceptors. Fig. 5.23 shows the estimated marginal means on a plot.

Using the SUS "acceptability ranges" [35] and "adjective ratings" [36], the overall average rating of SS was "good" and "acceptable"[38]. GP scored a little lower, with SUS "good" rating, but "marginal (high)" acceptance[39]. The rating for TS was "OK / fair", but it was "not acceptable" on the acceptance scale[40].

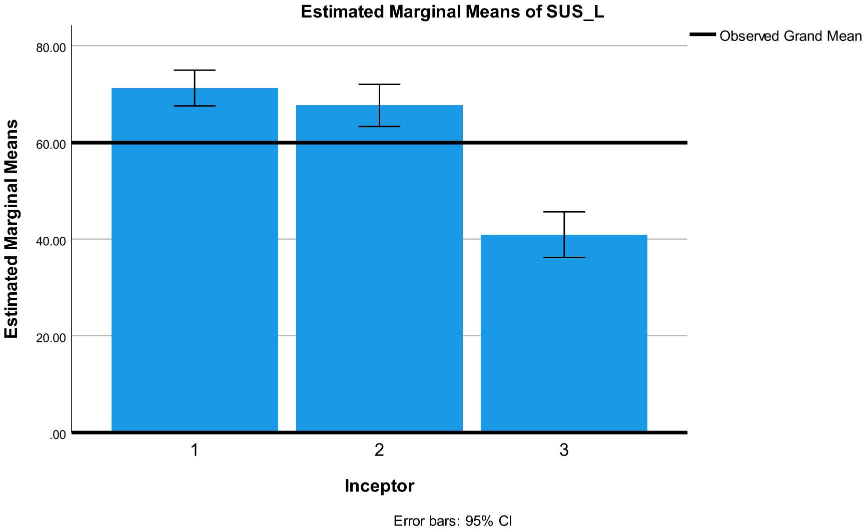


Figure 5.23: Estimated marginal means for SUS-Total. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.5 SART-D: Situational awareness rating technique - demand

The SART-D was the only SART subdimension that any grouping factors might have had an effect on. This section investigates if grouping factors GP usage and MG usage had a significant (and relevant) effect on the results (specifically GP usage on GP results and MG usage on TS results was of interest).

---

[38] With SD, SS ranged up to "best", and acceptance level was from "marginal (low)" to "acceptable".

[39] GP scales were ranging from "OK / fair" to "best", and included full spectrum of acceptance levels - "not acceptable" to "acceptable".

[40] With TS scores being "worst" to "good", and up to "marginal (low)" acceptance.

Gamepad usage as an independent factor

Grouping by GP usage was done by categorising participants by their frequency of using GP controller when playing VGs. There were four groups: "no", "hardly ever", "sometimes", and "a lot". Descriptive statistics are presented in Tab. 5.25, and Fig. 5.24. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 138) = 18.448$, $p < .001$, $\eta_p^2 = .211$) and the GP usage ($F(3, 69) = 2.933$, $p < .05$, $\eta_p^2 = .113$), but the interaction between them was not significant.

Although the main effect of GP usage was significant, pairwise comparisons and post-hoc tests did not reveal any significant differences between GP usage groups and interactions that would be relevant in this analysis. The comparisons of the main effect of the inceptor are investigated separately in the later section (page 143).

Data from the analysis can be found in Tab. A.54, A.55, A.56, A.57, and A.58, with bar plots of SART-D mean scores for each inceptor and GP usage shown in Fig. 5.24.

Results from this section show that GP usage did not have any significant effect on the results; therefore, it was excluded from further analyses.

Table 5.25: Descriptive statistics for SART-D scores with GP usage as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

GP usage	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no	10.29	3.24	10.68	3.90	13.67	3.57	27
yes - hardly ever	9.79	3.03	10.02	3.22	12.95	2.13	14
yes - sometimes	10.96	3.41	10.05	2.77	11.95	3.76	20
yes - a lot	12.92	2.76	11.90	2.04	14.83	2.17	12
Total	10.81	3.27	10.58	3.23	13.25	3.29	73

Mobile games usage as an independent factor

This section investigates the effect of MG usage on SART-D scores. Descriptive statistics are presented in Tab. 5.26, and Fig. 5.25. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. The rANOVA results showed

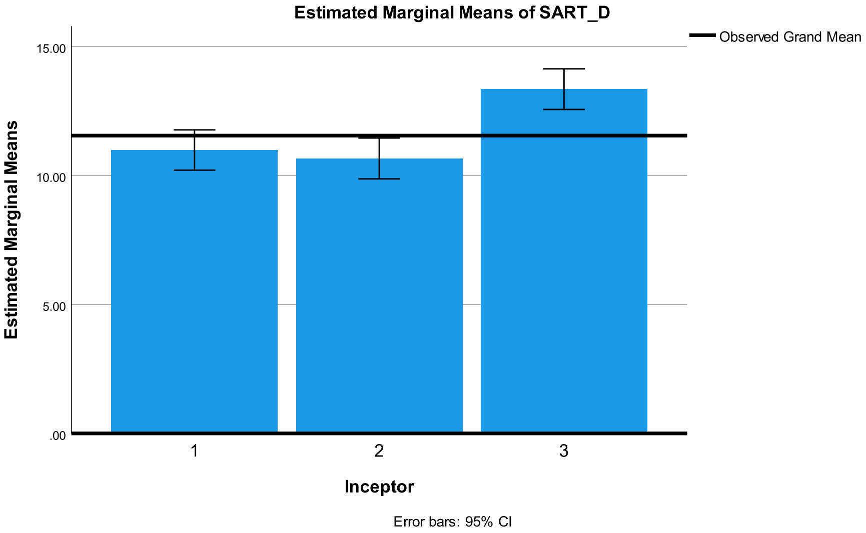


Figure 5.24: Estimated marginal means for SART-D within participant's GP usage. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

that there was a significant effect of inceptor: $F(2, 140) = 21.163$, $p < .001$, $\eta_p^2 = .232$, and of the inceptor * MG interaction: $F(4, 140) = 2.569$, $p < .05$, $\eta_p^2 = .068$. Between-subject effect of MG was not significant.

Pairwise comparisons and post-hoc tests did not show significant differences between MG usage groups. There was a significant difference between inceptors' main effect: the demand of SS and GP was rated significantly lower than TS's. A more detailed analysis of this observation can be found in a summary of this section (page 143). Inceptor * MG usage interaction analyses did not reveal significant differences in TS SART-D scores between participants with different MG usage. MG usage * inceptor analyses had similar patterns to the overall comparisons between inceptors: participants tended to score the demand of TS higher than SS and GP.

Data from the analysis can be found in Tab. A.59, A.60, A.61, A.62, and A.63, with bar plots of SART-D mean scores for each inceptor and MG usage shown in Fig. 5.25.

The analysis in this section did not confirm that the MG usage had a significant effect on the results, so this factor was excluded from further SART analyses.

Table 5.26: Descriptive statistics for SART-D scores with MG usage as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

MG uage	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no / hardly ever	10.04	3.40	11.55	3.22	13.53	3.39	29
used to	11.85	3.54	10.80	3.17	13.75	3.04	20
yes	10.86	2.74	9.22	2.93	12.49	3.37	24
Total	10.81	3.27	10.58	3.23	13.25	3.29	73

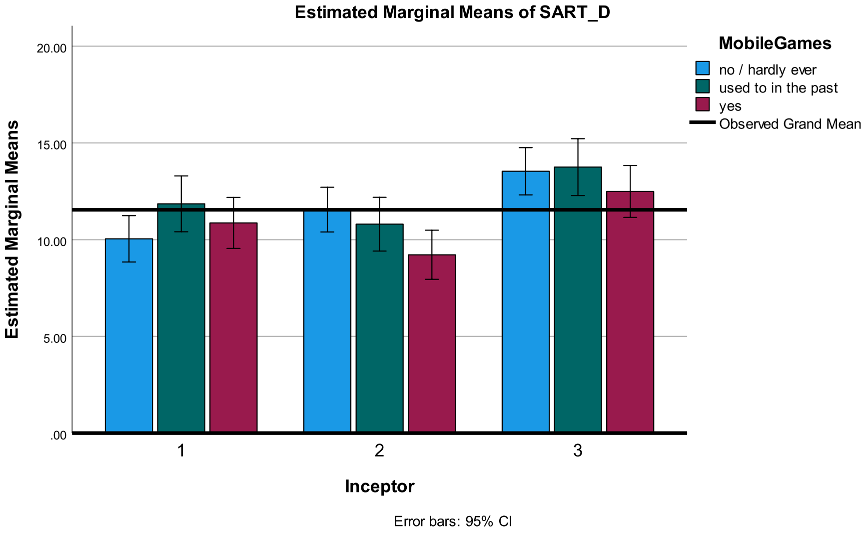


Figure 5.25: Estimated marginal means for SART-D across MG usage. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Summary

The analyses of GP and MG usage showed that those groups did not affect the SART-D results in this study. Therefore, a separate 1-way rANOVA was performed to confirm the main effect of the inceptor on the SART-D results. The visualisation of the scores is shown in Fig. 5.26.

The assumption of sphericity was not violated. There was a significant main effect of the inceptor: $F(2, 144) = 20.823$, $p < .001$, $\eta_p^2 = .224$. Estimated means of the inceptors were $M = 10.808$ for SS, $M = 10.579$ for GP, and $M = 13.250$ for TS.

Pairwise comparisons of estimated marginal means showed that SS and GP had approximately 2.5 points less than TS ($p < .001$). The difference between SS and GP was non-significant. Using the SART scale from Sec. 4.7.4, the demand of all three inceptors could be described as "moderate"[41].

Overall, this section could be concluded that none of the independent factors affected the perceived demand of the inceptors. The demand across all three inceptors was rated as "moderate"; however, TS results tend to be significantly higher point-wise.

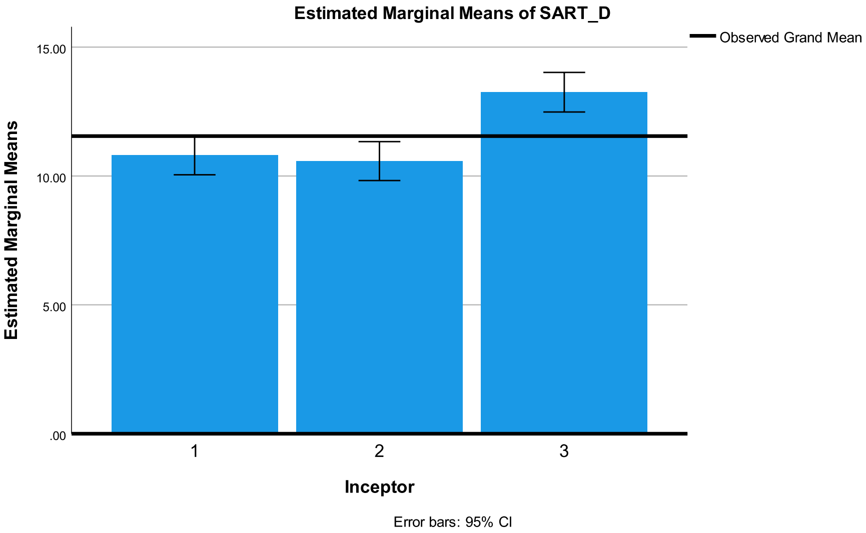


Figure 5.26: Estimated marginal means for SART-D. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.6 SART-S: Situational awareness rating technique - supply

SART-S scale did not have any significant interactions with the grouping factors, so the analysis of the results was based on a 1-way rANOVA. The assumption of sphericity was not violated. The within-subject (inceptor) effect was significant: $F(2, 144) = 3.614$, $p < .05$, $\eta_p^2 = .048$. Descriptive statistics are shown in Tab. 5.27, and the comparison is visualised as a bar graph in Fig. 5.27. Even though there was a potential (trending) effect of the inceptor effect in the main analysis, pairwise comparisons did

---

[41] With SD taken into consideration, ratings from SS and TS reached "low" and "high" demand, while ratings from GP were from "moderate" to "high".

not show any significant differences between the inceptors. Detailed results can be found in Tab. 5.28.

This means that participants felt that they were supplied with a similar amount of information for all of the inceptors. Based on the categorisation assumed in Section 4.7.4, the average supply level could be assumed as "moderate" (between 16 and 21) across all three inceptors. Putting aside the significance, participants even tended to give TS the highest SART-S score (approximately 1 point higher than both SS and GP).

Table 5.27: Descriptive statistics and estimates for SART-S scores. M - mean; SD - standard deviation; N - number of samples; SE - standard error; LB - lower bound; UB - upper bound. The value of M was the same for descriptive statistics and estimates. LB and UB are in a 95% confidence interval.

	Descriptives			Estimates		
	M	SD	N	SE	LB	UB
SS	18.09	3.32	73	.389	17.31	18.86
GP	18.05	3.11	73	.364	17.32	18.78
TS	19.10	3.44	73	.403	18.29	19.90

Table 5.28: Pairwise comparisons based on estimated marginal means for SART-S scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inceptor		MD	SE	p	LB	UB
SS	GP	.034	.416	1.000	-.985	1.054
	TS	-1.010	.461	.095	-2.139	.119
GP	SS	-.034	.416	1.000	-1.054	.985
	TS	-1.045	.447	.066	-2.139	.050
TS	SS	1.010	.461	.095	-.119	2.139
	GP	1.045	.447	.066	-.050	2.139

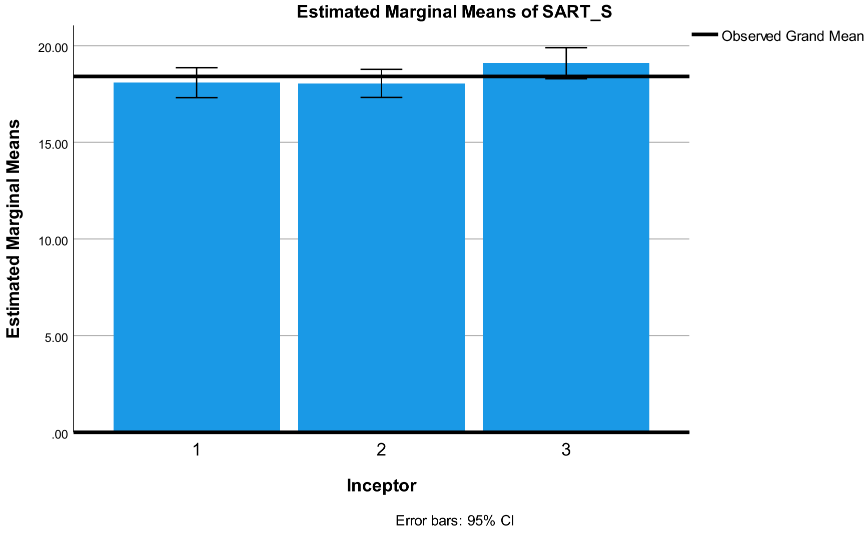


Figure 5.27: Estimated marginal means for SART-S. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.7 SART-U: Situational awareness rating technique - understanding

SART-U scale did not have any significant interactions with the grouping factors, so the analysis of the results was based on a 1-way rANOVA. The assumption of sphericity was not violated. The within-subject (inceptor) effect was significant: $F(2, 144) = 15.360$, $p < .001$, $\eta_p^2 = .176$. Descriptive statistics are shown in Tab. 5.29, and the comparison is visualised as a bar graph in Fig. 5.28.

Pairwise comparisons showed a significant difference between TS and SS (SS scored higher by 2.3 points, $p < .001$), and TS and GP (GP scored higher by 1.7 points, $p < .001$). SS and GP scores were not significantly different. Detailed results can be found in Tab. 5.30.

Although the scores between the inceptors were partly significant, the mean difference between them was not very high. The average understanding of all three inceptors could be categorised as "moderate" (SART-U scores between 12 and 16, according to the definition from Section 4.7.4). The lowest score for TS could be explained by the fact that this was a prototype of a new control method, which was unfamiliar to the participants.

Table 5.29: Descriptive statistics and estimates for SART-U scores. M - mean; SD - standard deviation; N - number of samples; SE - standard error; LB - lower bound; UB - upper bound. The value of M was the same for descriptive statistics and estimates. LB and UB are in a 95% confidence interval.

	Descriptives			Estimates		
	M	SD	N	SE	LB	UB
SS	15.31	2.91	73	.341	14.63	15.99
GP	14.69	3.57	73	.417	13.86	15.52
TS	13.00	3.99	73	.468	12.07	13.93

Table 5.30: Pairwise comparisons based on estimated marginal means for SART-U scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

		MD	SE	p	LB	UB
SS	GP	.616	.390	.356	-.340	1.573
	TS	2.308	.453	$< .001$	1.198	3.419
GP	SS	-.616	.390	.356	-1.573	.340
	TS	1.692	.448	.001	.594	2.789
TS	SS	-2.308	.453	$< .001$	-3.419	-1.198
	GP	-1.692	.448	.001	-2.789	-.594

5.7.8 SART-Total: Situational awareness rating technique - total score

SART-Total scale did not have any significant interactions with the grouping factors, so the analysis of the results was based on a 1-way rANOVA. The assumption of sphericity was not violated. The within-subject (inceptor) effect was significant: $F(2, 144) = 14.440$, $p < .001$, $\eta_p^2 = .167$. Descriptive statistics are shown in Tab. 5.31, and the comparison is visualised as a bar graph in Fig. 5.29.

Pairwise comparisons showed a significant difference between TS and SS (TS was scored 3.7 points lower, $p < .001$), and TS and GP (3.3 points less for TS, $p < .001$).

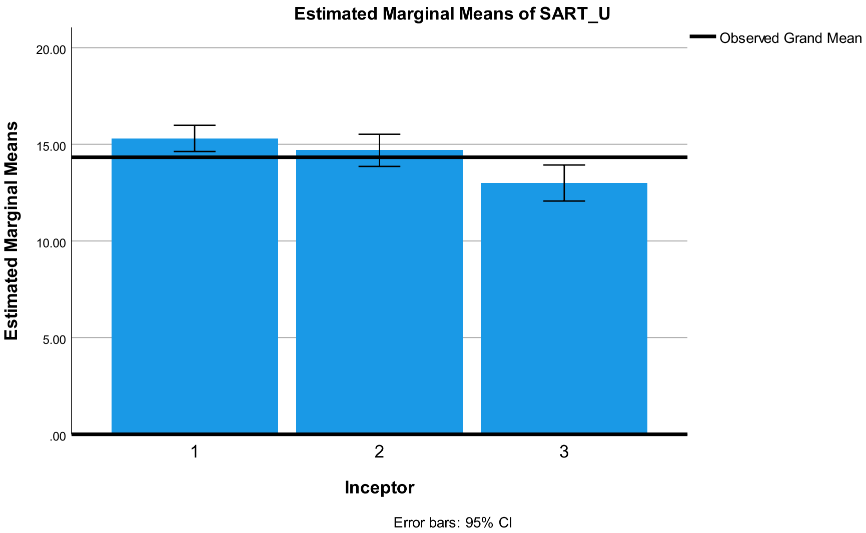


Figure 5.28: Estimated marginal means for SART-U. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

SS and GP scores were not significantly different. Detailed results can be found in Tab. 5.32.

Although the scores between the inceptors were partly significant, the mean difference between them was not very high. The average understanding of all three inceptors could be categorised as "moderate" (SART-Total scores between 12 and 16). The mean SART-Total score, indicating SA, was very similar for SS and GP. TS had significantly lower scores, with a mean difference of approximately 3.5 points less than SS and GP; nonetheless, the SA for all three inceptors could be classified as "moderate" (see Section 4.7.4).

Table 5.31: Descriptive statistics and estimates for SART-Total scores. M - mean; SD - standard deviation; N - number of samples; SE - standard error; LB - lower bound; UB - upper bound. The value of M was the same for descriptive statistics and estimates. LB and UB are in a 95% confidence interval.

	Descriptives			Estimates		
	M	SD	N	SE	LB	UB
SS	22.59	5.13	73	.601	21.39	23.78
GP	22.16	5.42	73	.634	20.90	23.43
TS	18.85	6.72	73	.787	17.28	20.41

Table 5.32: Pairwise comparisons based on estimated marginal means for SART-Total scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

		MD	SE	p	LB	UB
SS	GP	.421	.693	1.000	-1.277	2.120
	TS	3.740	.783	$< .001$	1.820	5.659
GP	SS	-.421	.693	1.000	-2.120	1.277
	TS	3.318	.806	$< .001$	1.342	5.295
TS	SS	-3.740	.783	$< .001$	-5.659	-1.820
	GP	-3.318	.806	$< .001$	-5.295	-1.342

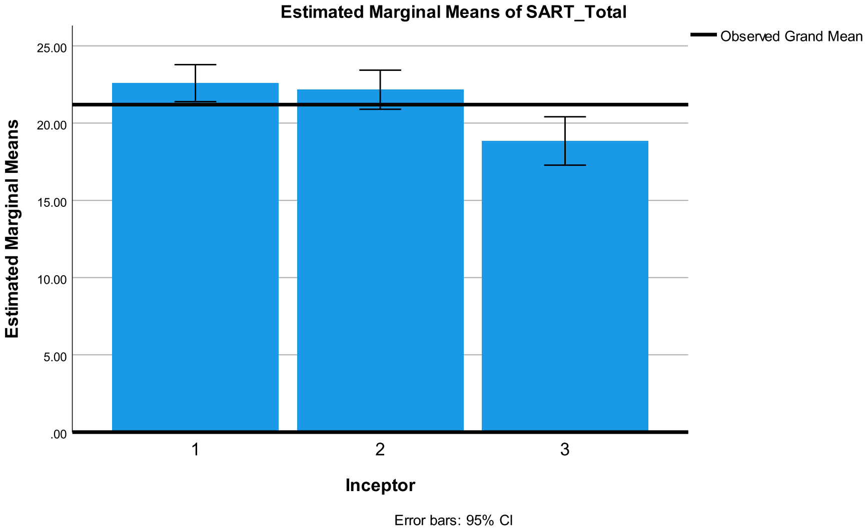


Figure 5.29: Estimated marginal means for SART-Total. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.9 NASA-TLX: Task Load Index

In the previous analysis in Section 5.6, it was found that only one grouping factor might have had a potential trending effect on the results - VG Group. To investigate this further, a 2-way mixed-design ANOVA was carried out. NASA-TLX scores across the three inceptors were a repeated-measures dependent factor, and VG Group (A, B, or C) was an independent factor. Descriptive statistics are presented in Tab. 5.33 and visualised as bar plots in Fig. 5.30 and 5.31. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. The main effect of inceptor was significant: $F(2, 138) = 28.526$, $p < .001$, $\eta_p^2 = .293$, but there was no interaction between the inceptor and VG group. Between-subject effect of VG group was not significant at 5% level. The p value was $p = .06$, so if the tolerance was elevated to 10%, the effect could be assumed to be showing a trend: $F(2, 69) = 2.926$, $p < .1$, $\eta_p^2 = .078$.

Pairwise comparisons between VG groups did not show significant differences; there was a trend that VG group C had approximately 9 points higher NASA-TLX score than group B (when the error tolerance was raised to 10% level, $p < .1$). Interestingly, VG group B (less frequent gamers) results showed the lowest perceived workload score: the mean NASA-TLX score for VG group B was $M = 47.338$, for VG group A - $M = 49.946$ (more frequent gamers), and for VG group C (non-gamers) - $M = 56.498$. Post-hoc analyses with Bonferroni correction confirmed the trend from pairwise comparisons by showing a significance that group B had a 9-point higher score than group C ($p < .05$). Analyses of the main effect of inceptor showed significant differences between all inceptors: SS was 5.6 points higher than GP ($p < .05$), TS was 10 points higher than SS ($p < .001$), and 16.5 points higher than GP ($p < .001$).

Pairwise comparisons of interaction showed a pattern within SS and GP between VG groups C and B: trending pattern of 12 points more in group C for SS ($p = .053$), and a significant pattern for GP - 14.6 points more in group C than B ($p < .05$). None of the pairwise comparisons of TS - VG group interaction showed a significant difference.

The interactions of inceptors within VG groups had similar patterns. In VG group A, the mean difference between TS and the other two controllers was $MD = 11.622$ for TS over SS ($p < .001$), and $MD = 17.463$ over GP ($p < .001$). In VG group B, TS scores were 16 points higher than SS ($p < .001$), and almost 23 points higher than GP. There were no significant differences between SS and GP in VG groups A and B nor within VG group C for any controller.

Detailed results can be found in Tab. A.64, A.65, A.66, and A.67, with bar plots of NASA-TLX mean scores for each inceptor shown in Fig. 5.30, and with separate VG groups in Fig. 5.31.

In summary, NASA-TLX results indicate that there were some significant differences in perceived workload between gamers and non-gamers when using SS and GP and between the three inceptors overall. However, the mean difference MD between the

SS and GP ($MD = 5.603$) was much less than TS compared to SS/GP ($MD = 10.926$ / $MD = 16.526$). This indicated that the GP was the lowest workload-inducing inceptor, followed by SS and TS. Within VG group C, each inceptor was approximately 5 points apart[42]. Given the 100-point NASA-TLX scale, such a low difference suggested that the workload difference between inceptors was insignificant. Expectedly, TS caused the most workload for participants due to the unfamiliarity and, in some cases, inverted Y-axis. In the case of TS (across all VG groups) and non-gamers (across all inceptors), the mean differences were not significant.

A possible reason for moderate gamers (group B) experiencing lower overall workload for SS and GP might be that the more frequent gamers (group A) felt the task was "more challenging"; thus, they tried to perform as well as they could. Given the high difficulty of the disturbance tasks, they might have felt more challenged and focused on the goal.

Table 5.33: Descriptive statistics for NASA-TLX scores with VG group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

VG Gr.	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	48.02	14.49	42.18	13.00	59.64	18.11	34
B	44.24	14.82	37.44	13.69	60.33	10.33	24
C	56.20	14.73	52.03	21.31	61.26	18.22	14
Total	48.35	15.04	42.52	15.78	60.19	15.75	72

---

[42] Mean difference between GP and SS, and between SS and TS was approximately 5 points on NASA-TLX scale.

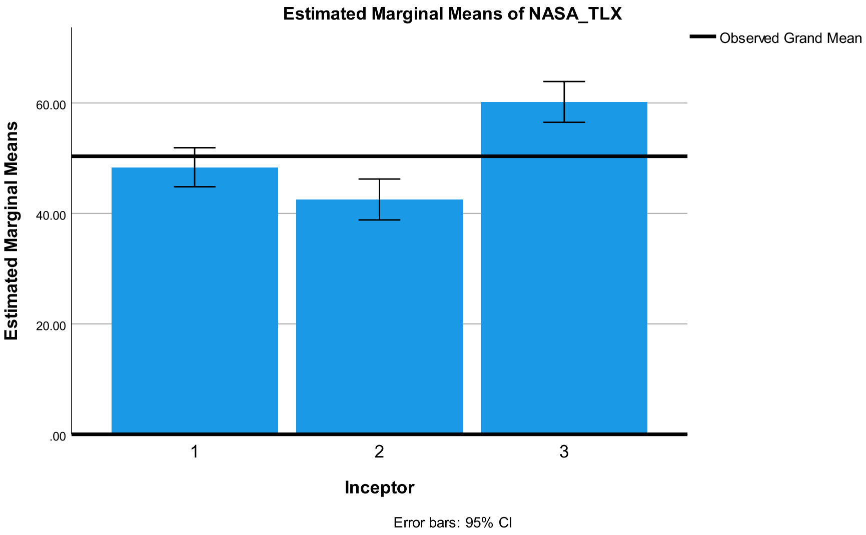


Figure 5.30: Estimated marginal means for NASA-TLX. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

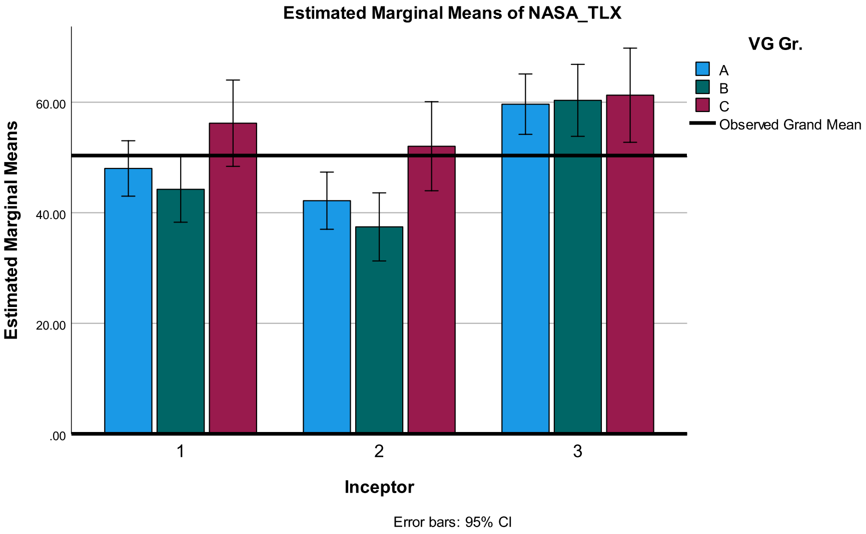


Figure 5.31: Estimated marginal means for NASA-TLX across VG Groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.10 Performance Score - DRV scenario

PS results from DRV scenario (along with LN) had the biggest number of potentially significant factors. Those factors were gender, hand, frequency of playing VG[43], MG,

---

[43] The "VG group" factor, however, appeared to be non-significant.

inceptor order, and FE. It was predicted that the significance of gender, hand and inceptor order was coincidental or irrelevant in this analysis. This section presents the results of 2-way mixed-design ANOVA, performed for each of the listed factors.

Gender as an independent factor

This section investigates the effect of gender on PS scores in DRV scenario. Descriptive statistics are presented in Tab. 5.34, and Fig. 5.32. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 10.882$, $p < .001$, $\eta_p^2 = .135$) and gender ($F(2, 70) = 6.824$, $p < .01$, $\eta_p^2 = .163$), but their interaction was not significant.

Pairwise comparisons showed a significant effect between male and female participants, and this pattern repeated in every inceptor. It was shown, however, that those differences might have resulted from gender distribution among participants with flight and VG experience (see sections 5.7.3, page 126 and 5.7.4, page 132), so it was decided to exclude this factor from further analyses, confirming the assumption from the beginning of Section 5.7.10. The main effect of the inceptor was analysed later in this section.

Data from the analysis can be found in Tab. A.69, A.70, A.71, and A.72.

Table 5.34: Descriptive statistics for PS in DRV scenario with gender as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Female	52.98	25.87	63.50	27.43	38.02	15.76	18
Male	70.46	19.70	77.08	16.35	53.14	17.24	54
Prefer not to say	88.03		95.93		56.39		1
Total	66.39	22.53	73.99	20.39	49.46	17.92	73

Handedness as an independent factor

This section investigates the effect of handedness on PS scores in DRV scenario. Descriptive statistics are presented in Tab. 5.35, and Fig. 5.33. Mauchly's test of

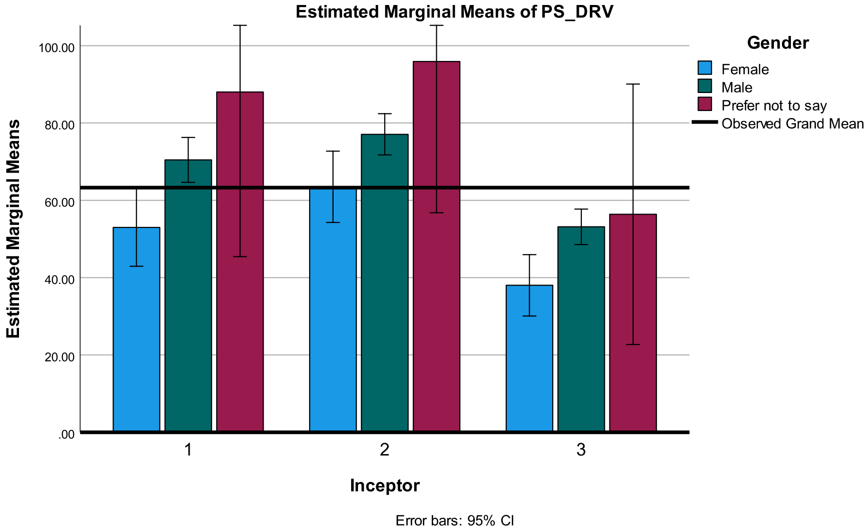


Figure 5.32: Estimated marginal means for PS DRV within participants' gender. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 14.435$, $p < .001$, $\eta_p^2 = .171$), but there was no significant main effect of handedness or inceptor * hand interaction.

Pairwise comparisons and post-hoc analysis did not show any significant effects or patterns other than that of the main inceptor effect. The pattern of GP scores being the highest, then SS and then TS appeared in every subgroup (although it was not significant everywhere, which could be caused by high inequality of sample sizes). Therefore, the assumption that the handedness factor was not significant (from the beginning of Section 5.7.10) was confirmed.

Data from the analysis can be found in Tab. A.73, A.74, A.75, A.76, and A.77.

Table 5.35: Descriptive statistics for PS in DRV scenario with handedness as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Ambidextrous	81.30	19.92	92.69	6.99	70.10	20.46	3
Left-handed	58.16	17.00	77.04	17.95	53.67	19.16	7
Right-handed	66.60	23.02	72.76	20.75	48.01	17.29	63

continued . . .

Table 5.35: … continued

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Total	66.39	22.53	73.99	20.39	49.46	17.92	73

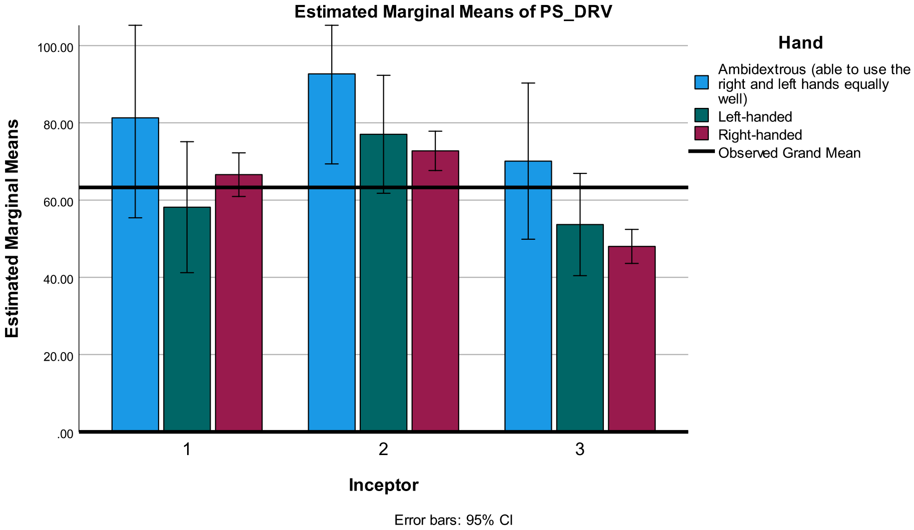


Figure 5.33: Estimated marginal means for PS DRV within participants' handedness. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Video game frequency as an independent factor

This section investigates the effect of VG frequency on PS scores in DRV scenario. Descriptive statistics are presented in Tab. 5.36, and Fig. 5.34. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 136) = 66.761$, $p < .001$, $\eta_p^2 = .495$); the VG frequency was showing a trend when the tolerance was elevated to 10% ($F(4, 68) = 2.467$, $p < .1$, $\eta_p^2 = .127$). The inceptor * VG frequency interaction was not significant.

Pairwise comparisons and post-hoc tests of the VG frequency's main effect were non-significant. The inceptor pattern was the same as in the previous section (page 153), and it was observable in most subgroups in pairwise comparisons of VG frequency * inceptor interaction.

Data from the analysis can be found in Tab. A.78, A.79, A.80, A.81, and A.82, with bar plots of PS mean scores for each inceptor and VG frequency in DRV scenario shown in Fig. 5.34.

Overall, it was found that VG frequency did not have a significant effect on the results. DRV was the only scenario where the initial analysis indicated potential VG influence on the PS; therefore, it could be assumed that gaming frequency did not have an impact on the inceptor performance.

Table 5.36: Descriptive statistics for PS in DRV scenario with VG frequency as a grouping factor. hpw - hours per week; M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no / hardly ever	64.21	20.55	74.19	21.31	42.36	20.53	14
used to, > 3 hpw	57.75	25.94	66.26	24.18	48.74	18.92	22
used to, < 3 hpw	61.58	23.65	69.51	16.94	47.22	13.19	11
yes, > 3 hpw	77.23	19.77	83.19	15.56	55.49	18.05	13
yes, < 3 hpw	76.58	12.24	81.45	14.76	54.18	15.64	13
Total	66.39	22.53	73.99	20.39	49.46	17.92	73

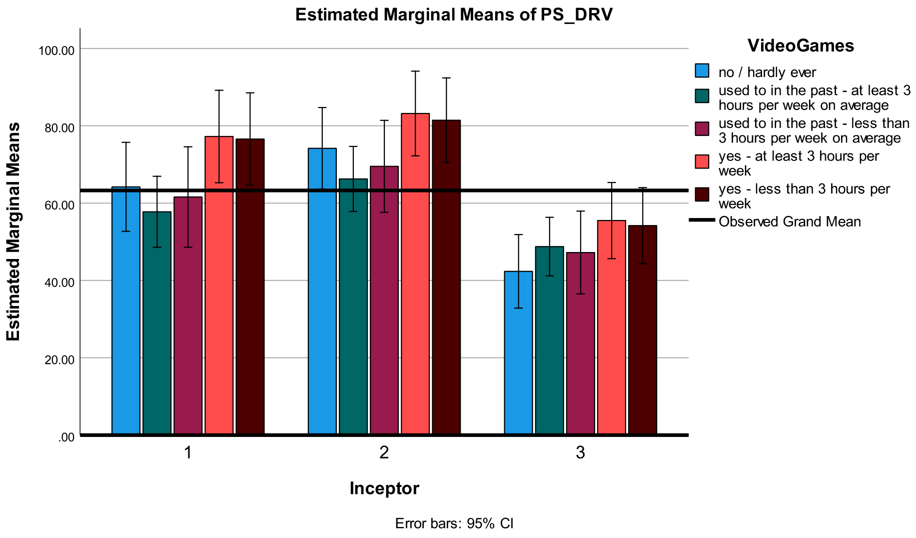


Figure 5.34: Estimated marginal means for PS DRV across VG groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Mobile games usage as an independent factor

This section investigates the effect of MG usage on PS scores in DRV scenario. Descriptive statistics are presented in Tab. 5.37 and Fig. 5.35. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. The rANOVA results showed that there was a significant effect of inceptor: $F(2, 140) = 61.624$, $p < .001$, $\eta_p^2 = .468$, but the inceptor * MG interaction and MG effects were not significant.

Pairwise comparisons did not reveal significant effects between MG groups, and interactions between inceptors within each group were convergent with the inceptor's effect.

Data from the analysis can be found in Tab. A.83, A.84, A.85, A.86, and A.87, with bar plots of PS mean scores for each inceptor and MG usage in DRV scenario shown in Fig. 5.35.

This analysis confirms that MG usage did not influence the PSs from DRV scenario, regardless of the used inceptor.

Table 5.37: Descriptive statistics for PS in DRV scenario with MG USAGE as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no / hardly ever	73.31	18.77	79.46	16.05	50.34	17.67	29
used to	59.45	24.27	70.39	20.63	48.74	17.54	20
yes	63.81	23.75	70.38	23.95	49.00	19.21	24
Total	66.39	22.53	73.99	20.39	49.46	17.92	73

Inceptor order as an independent factor

This analysis was performed to investigate if randomising the order of inceptors during the trials had an impact on the scores. Descriptive statistics are presented in Tab. 5.38, and Fig. 5.36. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. The rANOVA results showed that there is a significant effect of inceptor ($F(2, 134) = 71.378$, $p < .001$, $\eta_p^2 = .516$) and the inceptor * order interaction ($F(10, 134) = 2.236$, $p < .05$, $\eta_p^2 = .143$), but the order effects alone were

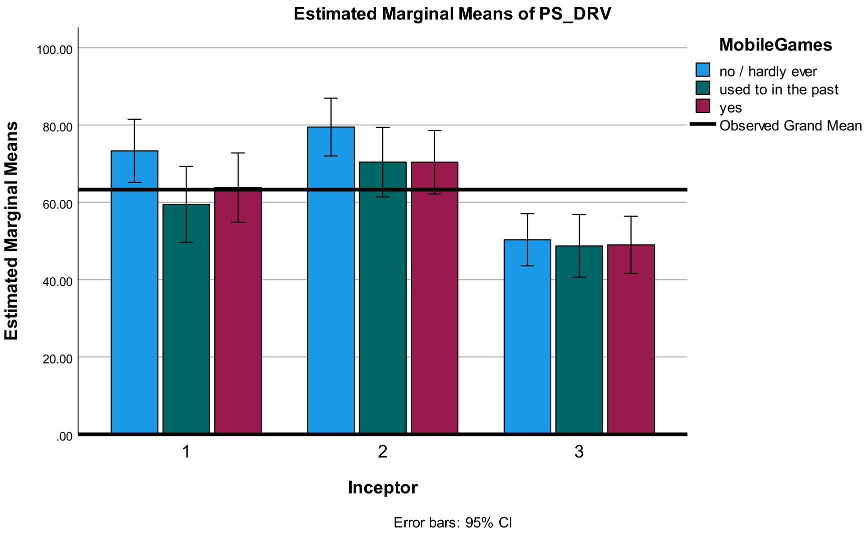


Figure 5.35: Estimated marginal means for PS DRV across MG usage. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

not significant.

Data from the analysis can be found in Tab. A.88, A.89, A.90, A.91, and A.92, with bar plots of PS in DRV scenario mean scores for each inceptor and inceptor order shown in Fig. 5.36.

Visual inspection of the scores showed that the last-used inceptor usually had slightly better scores as compared to other orders; the difference, however, was non-significant. Moreover, a small learning curve was expected, as with every attempt, participants were learning about the scenarios. Overall, the analysis confirmed that the inceptor order did not have a significant effect on the results, even though a small learning curve was observed.

Table 5.38: Descriptive statistics for PS in DRV scenario with Inceptor Order as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
123	56.86	26.89	74.28	18.32	52.25	18.61	13
132	61.24	21.97	74.65	20.01	38.77	18.02	12
213	64.96	23.60	67.45	25.07	54.79	18.79	13

continued . . .

Table 5.38: . . . continued

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
231	71.41	21.23	73.57	20.90	53.88	14.93	11
312	73.33	19.74	78.05	17.65	48.90	12.04	11
321	71.98	19.84	76.55	21.69	47.93	21.17	13
Total	66.39	22.53	73.99	20.39	49.46	17.92	73

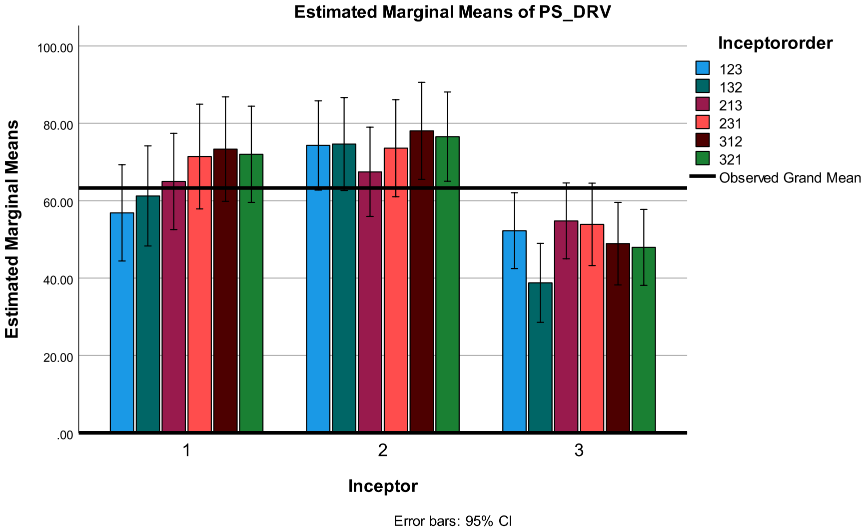


Figure 5.36: Estimated marginal means for PS DRV across inceptor order. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Flight experience group as an independent factor

This section investigates the effect of FE group on PS scores in DRV scenario. Descriptive statistics are presented in Tab. 5.39 and visualised in Fig. 5.37. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 75.925$, $p < .001$, $\eta_p^2 = .520$), FE group ($F(2, 70) = 14.034$, $p < .001$, $\eta_p^2 = .143$), and interaction between the inceptor and group ($F(4, 140) = 4.079$, $p < .01$, $\eta_p^2 = .104$).

Pairwise comparisons showed a significant difference between pilots and non-pilots: groups A and B had respectively mean PS of $M = 71.672$ and $M = 71.101$ (the difference between them was not significant); and group C had $M = 52.920$, differing significantly from group A and B ($p < .001$). Those observations were also confirmed in post-hoc analyses.

There were significant differences between all three inceptors ($p < .001$). SS scores were 7.7 points lower than GP's, but 18 points higher than TS's. The difference between GP and TS was almost 26 points.

Groups A and B did not have significant differences using SS; there was a significant difference in the performance of group C as compared to the other groups: they had 21 ($p < .01$) and 23 ($p < .001$) fewer points then groups A and B. GP results showed the same pattern: 26 ($p < .001$) and 17 ($p < .01$) fewer points then groups A and B. For TS, there was only one significant interaction, between groups C and B: group C had 14 fewer points ($p < .05$).

In group A, pairwise comparisons of FE group * inceptor interaction showed significant differences between all inceptors. Surprisingly, best scores were achieved using GP ($M = 88.485$), then SS ($M = 75.191$), and TS ($M = 51.339$). Group B did not have a significant difference between SS and GP, but the performance of both was significantly higher than TS. Mean score in group B was $M = 77.193$ for SS, $M = 79.032$ for GP, and $M = 57.077$ for TS. Group C had significant differences between all three inceptors, with GP scoring the highest, similar to group A. The mean performance was $M = 53.740$ for SS, $M = 61.918$ for GP, and $M = 43.102$ for TS.

Data from the analysis can be found in Tab. A.93, A.94, A.95, A.96, and A.97, with bar plots of PS DRV mean scores for each inceptor and FE group shown in Fig. 5.37.

In summary, FE showed to be the most significant factor on the performance in DRV scenario: pilots performed much better than non-pilots using all three inceptors; however, the amount of experience did not matter, as there was no significant difference between the high- and low-experienced pilots using any of the inceptors. Interestingly, GP was the best-performance inceptor, followed closely by SS. TS scores were the lowest overall, but pilots' performance using TS was comparable to non-pilots using SS and GP.

Table 5.39: Descriptive statistics for PS in DRV scenario with FE group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	75.19	20.05	88.49	8.78	51.34	20.20	19
B	77.19	19.38	79.03	16.25	57.08	18.70	22
C	53.74	19.89	61.92	21.09	43.10	13.67	32
Total	66.39	22.53	73.99	20.39	49.46	17.92	73

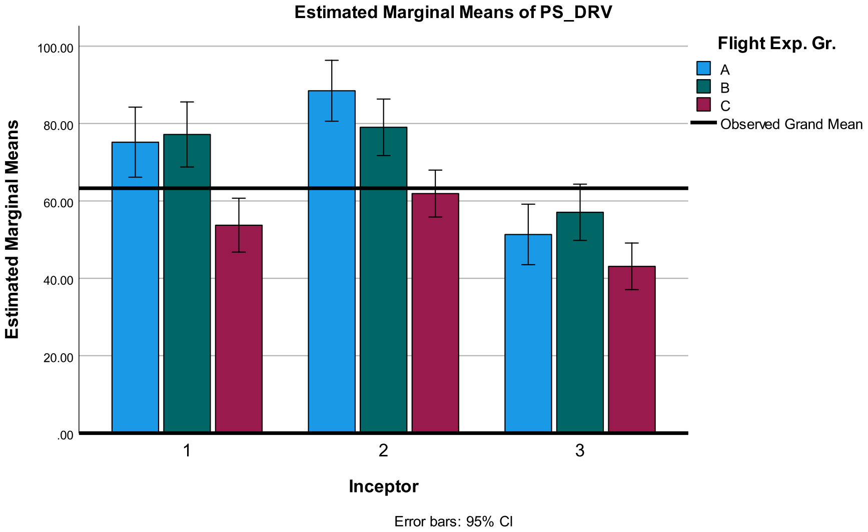


Figure 5.37: Estimated marginal means for PS DRV across FE groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Summary

Analyses in this section confirm the assumption from the beginning of Section 5.7.10 that gender, hand and inceptor order factors did not matter in the DRV scenario performance. Furthermore, it was found that the frequency of playing VG also was not significant. The only significant aspect was the FE. Pilots performed significantly better than non-pilots, albeit it did not matter whether they were professionals or beginners in aviation.

Overall scores showed that all participants performed best when using GP. Interestingly, the highest scores for that inceptor were achieved by high-experienced pilots on average ($M = 88.49$). This was followed by SS scores, where low-experienced pilots had slightly better performance than high-experienced pilots. TS scores were approximately in the middle of the PS scale ($M = 49.46$), which indicates that there is still much research to be done to investigate potential uses of this controller, but it does not have to be rejected, especially given that the difference value was less than one-fourth of the 100-point scale (average score of SS was $M = 66.39$, and GP - $M = 73.99$). Fig. 5.38 shows the mean score of each inceptor.

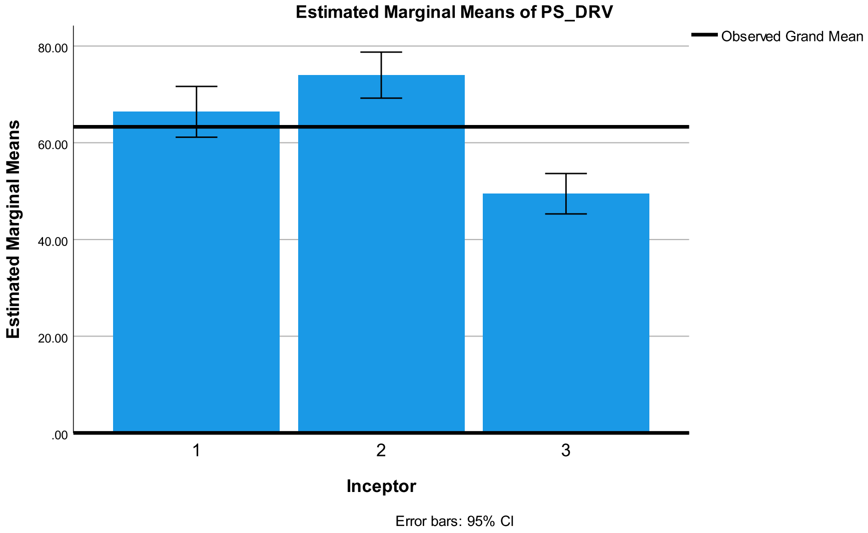


Figure 5.38: Estimated marginal means for PS DRV. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.11 Performance Score - DRH scenario

Initial analysis of PS from DRH scenario showed two potential factors influencing the results - gender and FE group. Based on the previous findings that gender did not influence the SUS and PS DRV results, it was decided to omit the detailed analysis of this factor and only report that the assumption of sphericity was not violated and the main effects of gender and inceptor * gender interaction were not significant. Detailed results of rANOVA with gender as a grouping factor can be found in Tab. A.99, A.100, A.101, and A.102.

The rest of this section presents the results from a 2-way mixed-design ANOVA, with FE group as an independent factor. Descriptive statistics are presented in Tab. 5.40

and visualised in Fig. 5.39. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 33.373$, $p < .001$, $\eta_p^2 = .323$) and FE group ($F(2, 70) = 4.095$, $p < .05$, $\eta_p^2 = .105$). The interaction between inceptor and group was not significant.

Pairwise comparisons showed that FE group A had the best performance (although non-significantly higher than group B's). Group B, in turn, did not differ significantly from group C. The only significant difference was between endpoint groups - A and C, with $MD = 14.606$ ($p < .05$), which was also the result of post-hoc tests. Inceptor differences were significant in every combination, with mean scores $M = 76.10$ for SS, $M = 88.34$ for GP, and $M = 71.39$ for TS, although the point difference was not large given a 100-point scale.

There were no significant interactions between FE groups among GP and TS results, and in SS, only groups A and C differed significantly ($MD = 19.615$, $p < .01$).

Investigation of each group separately revealed only one significant difference in group A: between GP and TS ($MD = 14.361$, $p < .001$); two differences in groups B and C: between SS and GP ($MD = -11.148$, $p < .05$ in B, and $MD = -16.137$, $p < .001$ in C), and between GP and TS ($MD = 18.071$, $p < .001$ in B, and $MD = 17.719$, $p < .001$ in C).

Data from the analysis with FE group as a grouping factor can be found in Tab. A.93, A.104, A.105, A.106, and A.107, with bar plots of SART-D mean scores for each inceptor and FE group shown in Fig. 5.39.

Table 5.40: Descriptive statistics for PS in DRH scenario with FE group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	88.10	8.30	95.04	7.54	80.68	14.27	19
B	76.81	24.24	87.96	18.96	69.89	22.07	22
C	68.48	26.01	84.62	19.57	66.90	23.92	32
Total	76.10	23.32	88.34	17.38	71.39	21.72	73

After the FE group analysis, a separate rANOVA was performed to confirm the main effect of the inceptor without any independent factors. Bar plots with estimated

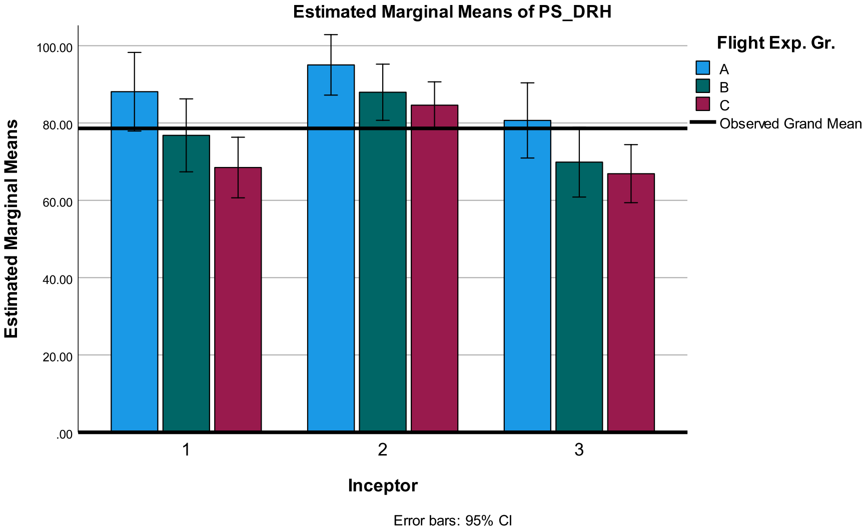


Figure 5.39: Estimated marginal means for PS DRH across FE groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

means are shown in Fig. 5.40. The results of the inceptor's main effect were similar to those from earlier sections - $F(2, 144) = 36.760$, $p < .001$, $\eta_p^2 = .338$; however, pairwise comparisons showed that there was a non-significant difference between SS and TS scores (there was a trend when the tolerance was elevated to 10%; the mean difference was $MD = 4.710$, $p < .1$).

Overall, the patterns in DRH PS were similar to DRV: GP scores were the highest, followed closely by SS, and then TS, visualised in Fig. 5.40. However, in this scenario, TS scores had a smaller mean difference to the other inceptors as compared to DRV, and even were non-significantly lower in 1-way rANOVA analysis. There was an evident influence of FE group on the scores, as seen in Fig. 5.39. FE group had smaller influence on DRH results when compared to DRV. The reason for that was/is? that DRH scenario was the easiest to perform.

5.7.12 Performance Score - LN scenario

PS results from LN scenario were thought to be influenced by the following factors: gender, handedness, GP usage, MG usage, inceptor order and FE group. Based on previous findings, it was highly possible that gender, handedness, and inceptor order would not have a significant effect. It was difficult to predict GP usage and MG usage results. FE group would most likely have had a significant impact. This section

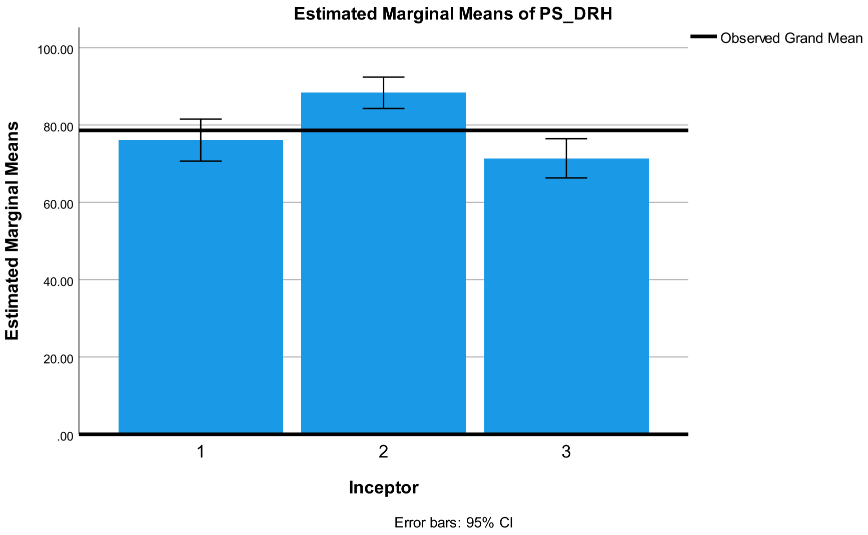


Figure 5.40: Estimated marginal means for PS DRH. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

investigates those hypotheses by performing 2-way mixed-design ANOVAs for each of the factors.

Gender as an independent factor

Similarly to previous analyses of gender factor (see Sections 5.7.10 and 5.7.11), it was found that it does not have a significant effect on the results. Results from rANOVA analysis showed that only the inceptor main effect was significant (the assumption of sphericity was not violated). Data from the analysis can be found in Tab. A.109, A.110, A.111, and A.112. This confirmed the assumption from the beginning of this section (5.7.12) that the gender factor would not be significant in the results of this study.

Handedness as an independent factor

This section investigates the effect of handedness on PS scores in LN scenario. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 4.276$, $p < .05$, $\eta_p^2 = .058$) and inceptor * hand interaction ($F(4, 140) = 2.581$, $p < .05$, $\eta_p^2 = .069$), but the effect size was negligible. There was no significant main effect of handedness. The post-hoc analysis did not reveal any significant comparisons.

Descriptive statistics and data from the analysis can be found in Tab. A.113, A.114, A.115, A.116, A.117, and A.118.

In the analysis of the handedness effect on PS DRV results (Section 5.7.10 on page 153), it was concluded that the sample size differences were too large to draw specific conclusions, so it was decided that this factor did not have a significant influence on the results, supporting the assumption from the beginning of this section (5.7.12).

Gamepad usage as an independent factor

This section investigates the effect of GP usage on PS scores in LN scenario. Descriptive statistics are presented in Tab. 5.41, and Fig. 5.41. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 138) = 13.855$, $p < .001$, $\eta_p^2 = .167$), but the GP usage and the interaction between them were not significant.

Pairwise comparisons and post-hoc tests did not reveal any significant differences between GP usage groups and interactions that would be relevant in this analysis, especially within GP results. There was a non-significant difference for participants who never used GP: they had 2-to-8-point lower mean score within GP results as compared to other participants; however, this might be connected with a VG group, as people who never used GP, naturally were also non-gamers. The comparisons of the main effect of the inceptor are investigated separately in the later section (page 172).

Data from the analysis can be found in Tab. A.119, A.120, A.121, A.122, and A.123, with bar plots of PS in LN scenario mean scores for each inceptor and GP usage shown in Fig. 5.41.

In summary, the amount of GP usage when playing VG did not significantly affect the scores.

Table 5.41: Descriptive statistics for PS in LN scenario with GP as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no	72.80	24.47	69.15	23.47	56.27	19.55	27
yes - hardly ever	67.24	30.38	74.49	25.44	61.07	26.23	14

continued ...

Table 5.41: . . . continued

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
yes - sometimes	76.05	24.52	71.19	22.03	60.73	19.98	20
yes - a lot	78.77	14.16	77.18	13.93	68.53	13.14	12
Total	73.60	24.22	72.05	21.96	60.43	20.29	73

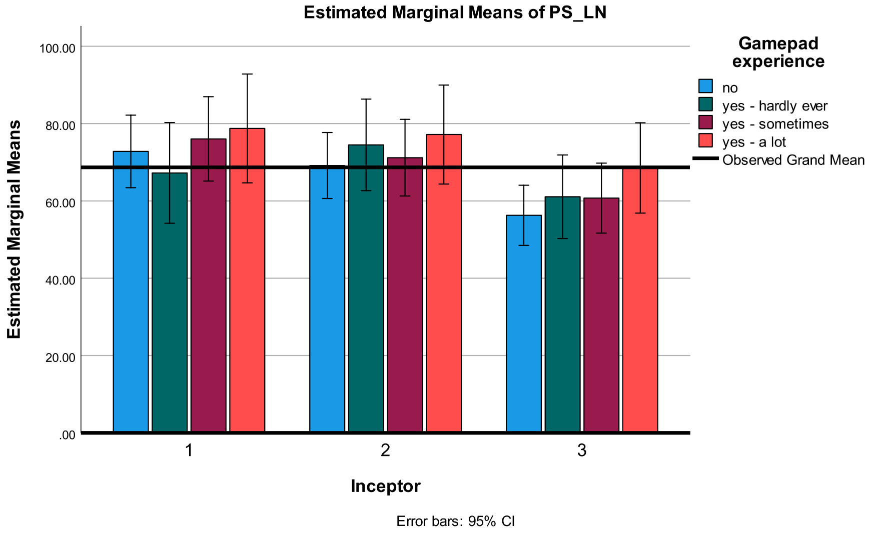


Figure 5.41: Estimated marginal means for PS LN across GP usage. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Mobile games usage as an independent factor

This section investigates the effect of MG usage on PS scores in LN scenario. Descriptive statistics are presented in Tab. 5.42, and Fig. 5.42. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. The rANOVA results did show that there is a significant effect of inceptor ($F(2, 140) = 17.474$, $p < .001$, $\eta_p^2 = .200$) and its interaction with MG ($F(2, 140) = 6.398$, $p < .001$, $\eta_p^2 = .155$), but the main effect of MG factor was not significant.

Pairwise comparisons and post-hoc tests did not reveal significant differences between MG groups and interactions between groups (within TS results). The comparisons of the main effect of the inceptor are investigated separately in the later section (page 172).

Data from the analysis can be found in Tab. A.124, A.125, A.126, A.127, and A.128, with bar plots of PS in LN scenario mean scores for each inceptor and MG usage shown in Fig. 5.42.

This analysis confirmed that MG usage did not influence the PSs from the LN scenario, regardless of the used inceptor.

Table 5.42: Descriptive statistics for PS in LN scenario with MG usage as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no / hardly ever	84.07	13.15	76.52	14.34	61.78	17.96	29
used to	59.88	28.97	74.53	22.08	58.81	24.94	20
yes	72.39	25.07	64.59	27.72	60.14	19.42	24
Total	73.60	24.22	72.05	21.96	60.43	20.29	73

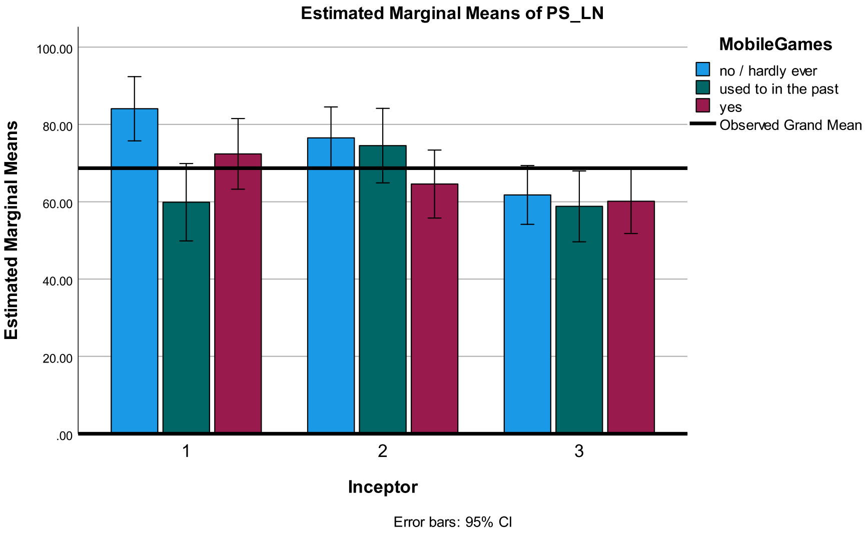


Figure 5.42: Estimated marginal means for PS LN across MG usage. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Inceptor order as an independent factor

This section investigates the effect of inceptor order on PS scores in LN scenario. Descriptive statistics are presented in Tab. 5.43, and Fig. 5.43. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. The rANOVA results showed that there was a significant effect of inceptor ($F(2, 134) = 19.906$, $p < .001$, $\eta_p^2 = .229$) and the interaction ($F(10, 134) = 2.427$, $p < .05$, $\eta_p^2 = .153$), but the inceptor order was not significant.

Pairwise comparisons and post-hoc tests did not reveal significant interactions between the inceptor order. For this scenario, only a non-significant learning curve could be observed for GP, where participants who tested this inceptor as a last one had better scores than other groups.

Data from the analysis can be found in Tab. A.129, A.130, A.131, A.132, and A.133, with bar plots of PS in LN scenario mean scores for each inceptor and inceptor order shown in Fig. 5.43.

In summary, the analysis confirmed that the inceptor order did not have a significant effect on the results. No learning curve was observed other than for the GP controller (non-significant). This was in contrast to the inceptor order effect on the PS DRV results, where a learning curve was slightly more evident. The reason for this is that the scenario order was fixed, so participants already familiarised themselves with the simulator. This showed that the learnability of the system was "quick" and easy.

Table 5.43: Descriptive statistics for PS in LN scenario with Inceptor Order as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
123	60.39	28.86	63.42	31.04	57.27	22.04	13
132	64.73	19.08	79.96	5.95	60.82	14.61	12
213	74.11	23.69	67.79	25.85	60.52	21.60	13
231	85.30	11.81	65.50	15.56	58.40	20.01	11
312	84.09	15.63	84.14	9.05	66.88	22.88	11
321	75.74	31.17	72.96	24.96	59.37	22.29	13
Total	73.60	24.22	72.05	21.96	60.43	20.29	73

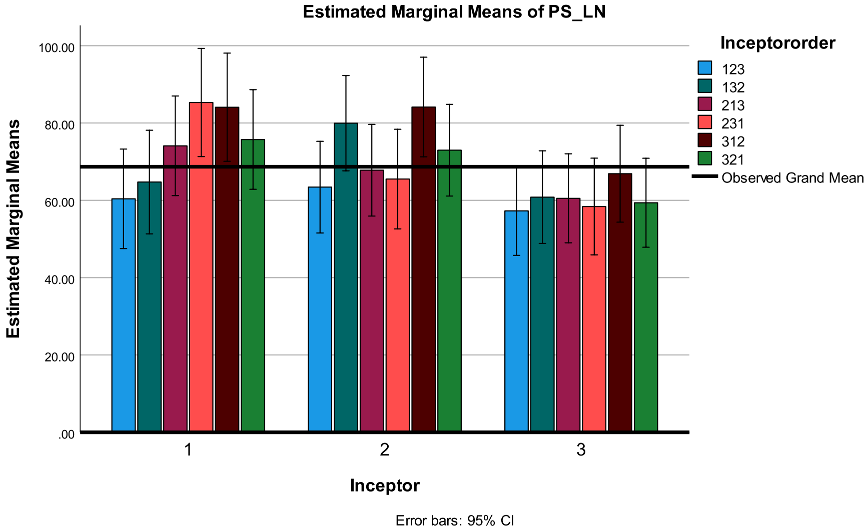


Figure 5.43: Estimated marginal means for PS LN across inceptor order. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Flight experience group as an independent factor

This section investigates the effect of FE group on PS scores in LN scenario. Descriptive statistics are presented in Tab. 5.44 and visualised in Fig. 5.45. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 22.913$, $p < .001$, $\eta_p^2 = .247$), FE group ($F(2, 70) = 5.695$, $p < .01$, $\eta_p^2 = .140$), and the interaction between inceptor and group ($F(4, 140) = 3.576$, $p < .01$, $\eta_p^2 = .093$).

It could be observed from pairwise comparisons and post-hoc results that FE groups A ($M = 76.301$) and B ($M = 73.414$) did not differ from each other, but they had significantly higher ($p < .05$) PS than group C ($M = 60.930$).

Inceptor analyses showed that SS and GP scores are the highest, with mean estimates $M = 75.869$ for SS, $M = 73.907$ for GP, and $M = 60.870$ for TS. The mean difference between SS and GP was not significant, but both these inceptors were significantly higher ($p < .001$) than TS.

Interaction analyses showed the biggest mean differences within SS results: between FE group A had 24 more points than C ($p < .01$), and group B had almost 16 points more than C ($p < .05$). Expectedly, there was no significant difference between groups A and B, although on average, group A had an approximately 8-point advantage over group B. GP had similar patterns to SS: there was no significant difference between groups A and B, and both had a significant 18- ($p < .01$) and 16-point ($p < .05$) advantage over group C, respectively. Interestingly, there were no significant differences among TS results, with groups A and B showing only a 4-point advantage over group C.

Within groups A and B, a similar pattern could be found. There were significant PS differences between SS and TS: group A had almost 25-point SS advantage over TS ($p < .001$), while group B's points advantage was 16 ($p < .01$). Between GP and TS, group A had an 18-point advantage ($p < .001$), and group B had 16 points more ($p < .01$). SS and GP showed no significant differences. There were no significant differences among group C, although the TS results were approximately 4 points lower than SS and GP.

Data from the analysis can be found in Tab. A.134, A.135, A.136, A.137, and A.138, with bar plots of PS LN mean scores for each inceptor are shown in Fig. 5.44, and across FE groups shown in Fig. 5.45.

In summary, patterns in PS LN results were slightly different from those in DR scenarios. Here, high-experienced pilots performed best using a SS (although the difference between GP results was non-significant), low-experienced pilots had very similar results when using SS and GP, and non-pilots had no significant differences in PS, regardless of the inceptor used. Interestingly, all groups achieved similar scores using the TS.

Table 5.44: Descriptive statistics for PS in LN scenario with FE group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	86.61	12.39	80.31	12.30	61.99	15.47	19
B	78.46	24.31	79.13	12.77	62.66	24.75	22
C	62.54	25.02	62.29	27.32	57.96	19.76	32
Total	73.60	24.22	72.05	21.96	60.43	20.29	73

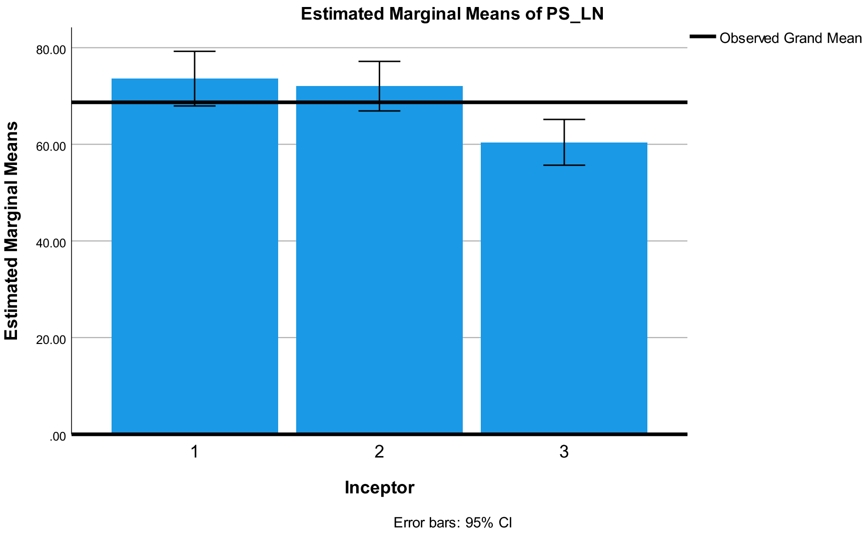


Figure 5.44: Estimated marginal means for PS LN. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Summary

Results from this section show that gender, handedness, and inceptor order did not significantly influence the PS from the LN scenario, confirming the assumptions from the beginning of this section (5.7.12, page 164). GP and MG usage groups were checked against the performance measures and also did not show significant effects. As expected, FE had a strong effect on the results.

In summary, overall results from this scenario showed the superiority of SS. However, the GP was slightly worse and only within the group of highly experienced pilots. The SS score differences in groups B and C were negligible. The average mean difference between pilots was approximately 18.8 points on the PS scale. Interestingly, non-pilots had almost comparable results, regardless of the controller; they achieved around 60 points on a PS scale on average, which was similar to pilots' performance with TS. The TS was an unknown controller for each group, and yet, they managed to accomplish satisfactory results. Given that it was the first experience with a flight simulation for many non-pilot participants, they managed to benefit from every inceptor and gain reasonable PS compared to pilots' results.

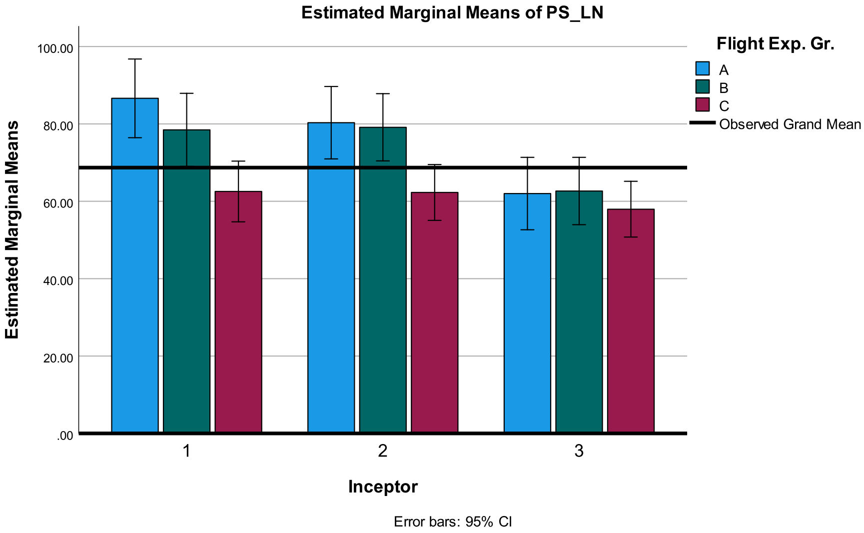


Figure 5.45: Estimated marginal means for PS LN across FE groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

5.7.13 Performance Score - LD scenario

Three factors were chosen for further investigation with regard to the LD scenario: gender, MG usage and FE. Based on previous findings, it was expected that the first two would not have a significant influence on the performance in this scenario, and FE would show a major impact.

Gender as an independent factor

This section investigates the effect of gender on PS scores in LD scenario. The results of rANOVA showed that, in this case, the inceptor and the interaction were not significant. There was a significant between-subject effect of gender ($F(2, 70) = 5.388$, $p < .01$, $\eta_p^2 = .133$). The assumption of sphericity was not violated. Data from the analysis can be found in Tab. A.140, A.141, A.142, and A.143. Further analysis revealed that the significant differences were only within SS and GP inceptors. Moreover, there were no significant interactions among female participants. The difference seems to be coincidental, and it appears that gender was a hidden factor of the FE group, as explained in Section 5.7.3 (page 126). This supports the assumption that the gender factor was not significant in this study.

Mobile games usage as an independent factor

This section investigates the effect of MG usage on PS scores in LD scenario. The assumption of sphericity was not met ($p < .05$), and the sphericity estimate was equal to Mauchly's $W = .893$. According to Girden, if the estimate is greater than 0.75, Huynh–Feldt correction is recommended [140]. Therefore, the results showed a significant main effect of inceptor ($F(1.904, 133.284) = 19.063$, $p < .001$, $\eta_p^2 = .214$) and the interaction with MG ($F(3.808, 133.284) = 3.057$, $p < .05$, $\eta_p^2 = .080$). The main effect of MG factor was not significant.

Pairwise comparisons and post-hoc tests did not reveal significant differences between MG groups and interactions between groups within TS results. Some significant relationships were found within the SS results, but they were assumed to be coincidental; higher SS results among participants that have not or hardly ever played MG could be explained by the high percentage of non-MG users among the high-experienced pilots, as seen in Fig. 5.46. The comparisons of the main effect of the inceptor are investigated in the next section.

Descriptive statistics and data from the analysis can be found in Tab. 5.45, A.144, A.145, A.146, A.147, and A.148.

This analysis confirmed that MG usage did not have a significant impact on the PSs from the LD scenario, which confirms the assumption from the beginning of this Section (5.7.13, page 173).

Table 5.45: Descriptive statistics for PS in LD scenario with MG usage as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
no / hardly ever	76.79	15.29	69.47	22.06	57.79	16.93	29
used to	58.33	23.23	71.64	23.95	50.43	26.90	20
yes	58.42	25.08	63.23	27.90	50.59	22.86	24
Total	65.69	22.73	68.01	24.52	53.41	21.96	73

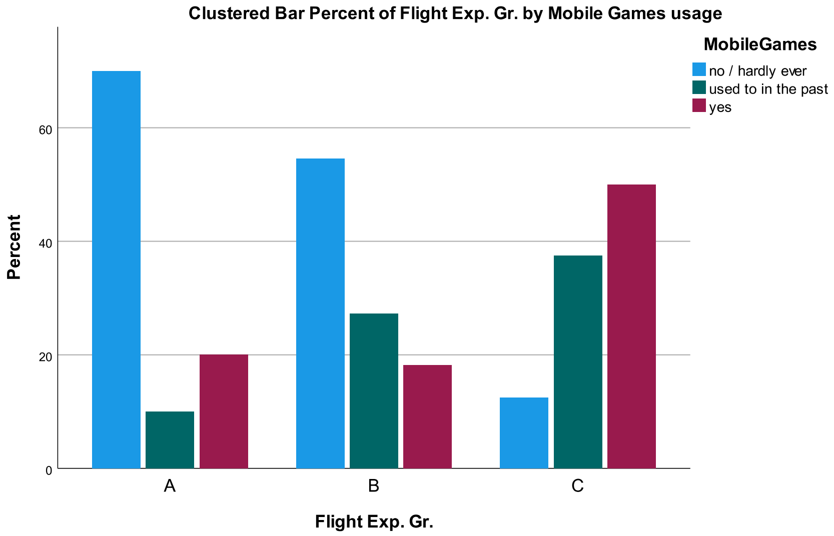


Figure 5.46: Distribution of MG usage across each FE group. Percentage within each group cluster sums up to 100%.

Flight experience group as an independent factor

This section investigates the effect of FE group on PS scores in LD scenario. Descriptive statistics are presented in Tab. 5.46 and visualised in Fig. 5.48. Mauchly's test of sphericity confirmed that the assumption of sphericity was not violated. There was a significant main effect of inceptor ($F(2, 140) = 20.880$, $p < .001$, $\eta_p^2 = .230$), FE group ($F(2, 70) = 10.191$, $p < .001$, $\eta_p^2 = .226$), and the interaction between inceptor and group ($F(4, 140) = 2.510$, $p < .05$, $\eta_p^2 = .067$).

Estimates within each FE group were $M = 73.363$ for group A, $M = 67.289$ for group B, and $M = 52.465$ for group C. The difference between groups A and B was observable but not significant, whereas group C differed significantly from the first two: the scores were 20 points lower than group A ($p < .001$) and 14 points lower than group B ($p < .01$). This was confirmed with post-hoc tests.

Similar to LN results, in LD scenario, the SS and GP results were the highest, not significantly different from each other, but both had significantly higher scores than TS by approximately 13-15 points ($p < .001$).

Among each inceptor in inceptor * FE group comparisons, the groups had similar patterns, and even MD values[44], as those from FE group analysis in the LN scenario (Section 5.7.12, page 170).

FE group * inceptor analyses also shared the pattern with LN results, with SS and GP results being significantly higher than TS, but not much different from each other[45].

Data from the analysis can be found in Tab. A.149, A.150, A.151, A.152, and A.153, with bar plots of PS LD mean scores for each inceptor are shown in Fig. 5.47, and across FE groups shown in Fig. 5.48.

In summary, the SS and GP results of FE group and inceptor effects and interactions from LD scenario are convergent with those from LN scenario (Section 5.7.12, page 170). High-experienced pilots achieved the best results with SS (GP scores were non-significantly lower); low-experienced pilots scored best using GP, but only a little difference to SS; however, the pattern in TS results was, albeit non-significantly, different. It was evident that in the turbulent scenario, FE matters. This can be seen in Fig. 5.48, where, among TS results, high-experienced pilots performed best, followed by their lower-experienced colleagues (with an 8-point difference), and non-pilots (with 14 points difference). The scores from non-pilots were the lowest because they were not accustomed to experiencing such turbulence in any situation, whereas pilots could have been trained in such conditions or even experienced them in real life.

Table 5.46: Descriptive statistics for PS in LD scenario with FE group as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
A	81.32	15.89	76.50	17.26	62.27	14.59	19
B	70.92	19.24	77.11	21.21	53.83	24.37	22
C	52.82	21.50	56.72	26.16	47.86	22.67	32

continued . . .

---

[44] Significant interactions of PS in LD scenario were: for SS, $MD = 28.501$ ($p < .001$) between A and C, and $MD = 18.104$ ($p < .01$) between B and C; for GP, $MD = 19.782$ ($p < .05$) between A and C, and $MD = 20.397$ ($p < .01$) between B and C; for TS, the differences where not significant.

[45] Significant interactions of PS in LD scenario were: for group A, $MD = 19.052$ ($p < .001$) between SS and TS, and $MD = 14.230$ ($p < .05$) between GP and TS; for group B, $MD = 17.092$ ($p < .001$) between SS and TS, and $MD = 23.282$ ($p < .001$) between GP and TS; for group C, the differences where not significant.

Table 5.46: ...continued

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Total	65.69	22.73	68.01	24.52	53.41	21.96	73

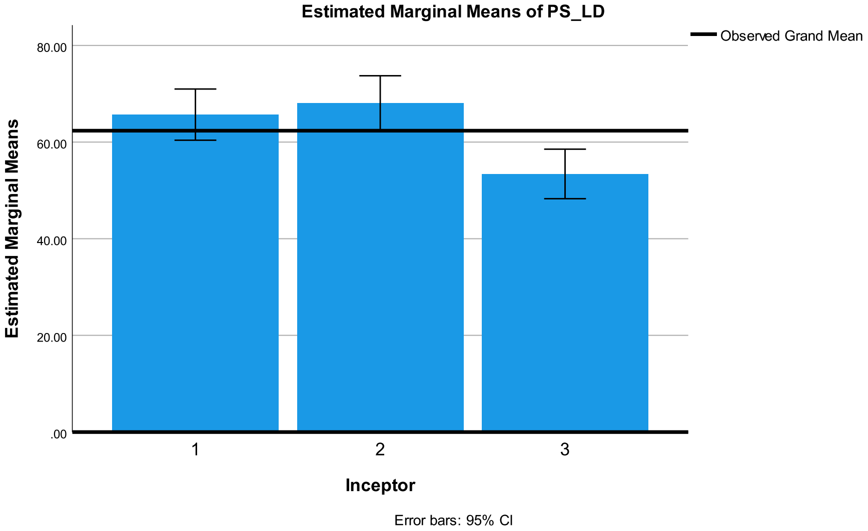


Figure 5.47: Estimated marginal means for PS LD. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

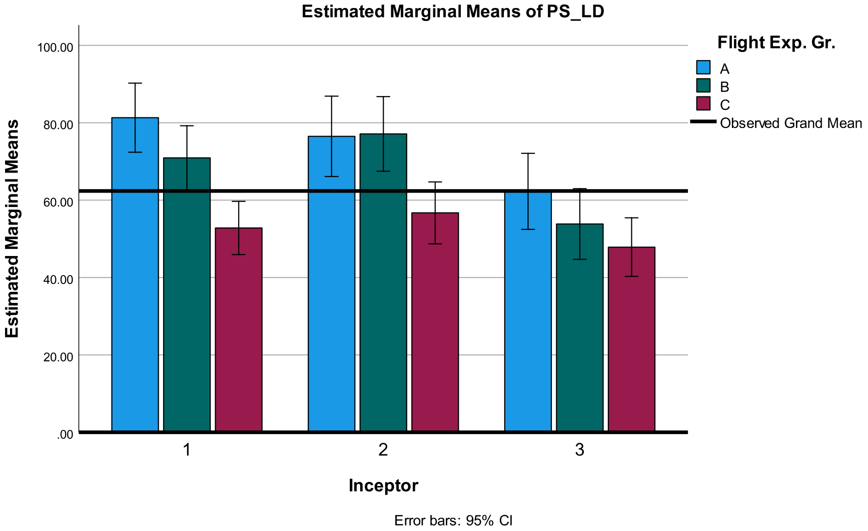


Figure 5.48: Estimated marginal means for PS LD across FE groups. Inceptors are coded as follows: 1 - SS; 2 - GP; 3 - TS.

Summary

Results from this section showed that gender and MG usage did not significantly influence the PS, which confirms the assumptions from the beginning of this section (5.7.13, page 173). FE appeared to be crucial in performing a landing in such difficult conditions.

It is evident that, in turbulence, SS was the safest inceptor among the three investigated in this study, but the GP was only slightly worse overall. It was found that groups A and B, and inceptors SS and GP, showed similar patterns to the respective results from LN (Section 5.7.12). It was found that non-pilots struggled more in turbulent conditions, and the reason might be a lack of experience.

Comparing those results with LN scenario revealed one interesting matter: for SS and GP, the PS decreased when the disturbance was added; so was the performance within low-experienced and naïve pilots for TS. However, the performance of high-experienced pilots did not significantly decrease but slightly increased instead. Because of this, further analyses were performed, where an additional factor was added to the rANOVA design - Landing scenario type. Results are presented in the next section.

5.7.14 Performance Score - LN and LD scenario interaction

In order to investigate the differences in PS between the LN and LD scenarios, three separate rANOVAs were carried out (for each inceptor), where a scenario type was a dependent factor (LN and LD), and the FE group was an independent factor. Since the main effect of the FE group was already investigated, only the results of the scenario were reported. Mauchly's test was not performed because the measured variable had two levels. Results are presented in Tab. 5.47, 5.48 and Fig. 5.49-5.51

In SS, the landing factor had a significant effect: $F(1,70) = 11.698$, $p < .01$, $\eta_p^2 = .143$. On average, all participants achieved 7.5 points more in the LN scenario, as compared to LD ($p < .01$). Within FE groups, those differences ranged from non-significant in group A, and trending in group B ($MD = 7.535$, $p < .1$[46], to a significant ($MD = 9.725$, $p < .01$) in group C.

In GP, there was no significant main effect of landing; only a trend when the tolerance was elevated to 10%: $F(1,71) = 3.400$, $p < .1$, $\eta_p^2 = .046$. The average score

---

[46] When the tolerance was elevated to 10%.

varied between the LN and LD by 4.2 points (LN scores were higher; $p < .1$). Pairwise comparisons did not show any significant interactions but revealed that groups A and C had a similar 5-point difference between LN and LD scores. Group B had almost the same result in both scenarios, so the difference did not show as significant.

The scenario type in TS had a significant difference on TS results: $F(1, 71) = 7.312$, $p < .01$, $\eta_p^2 = .093$, with LN scores being 6.1 points higher than LD ($p < .01$). Results from group A did not differ significantly and were even slightly higher in LD scenario on average (which can be seen in Fig. 5.51), while the performance decrease from LN to LD among groups B and C was significant ($MD = 8.828$, $p < .05$ in group B, and $MD = 10.102$, $p < .001$ in group C).

It is worth noting that the LD scenario was much more difficult than LN; nonetheless, participants' PS mean difference was not high, averaging at approximately 6 points on a 100-point scale. It shows that all three inceptors allowed participants to quickly get used to the given task and not lose focus. Interestingly, the decrease in performance within all FE groups was the lowest for GP. Perhaps the most interesting observation is that high-experienced pilots achieved the same score in both scenarios using TS, despite the significantly greater difficulty of the task. This proves that pilot training and substantial flying experience help pilots quickly adapt to novel prototype inceptors. While the TS inceptor is not necessarily applicable on a flight deck in its current form, there is a potential for introducing it in other areas, such as urban air mobility.

Table 5.47: Estimates for PS in LN-LD scenarios with FE group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval.

Inc.	Scenario	M	SE	LB	UB
SS	LN	75.87	2.664	70.56	81.18
	LD	68.35	2.339	63.69	73.02
GP	LN	74.07	2.413	69.26	78.88
	LD	69.86	2.681	64.51	75.20
TS	LN	60.70	2.413	55.89	65.52
	LD	54.57	2.529	49.52	59.61

Table 5.48: Pairwise comparisons of mean difference between LN and LD scenarios, based on estimated marginal means for PS in each inceptor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. Column MD shows an interaction between the scenarios (LN * LD).

Inc.	MD	SE	p	LB	UB
SS	7.515	2.197	.001	3.133	11.897
GP	4.211	2.284	.069	-.343	8.764
TS	6.136	2.269	.009	1.612	10.661

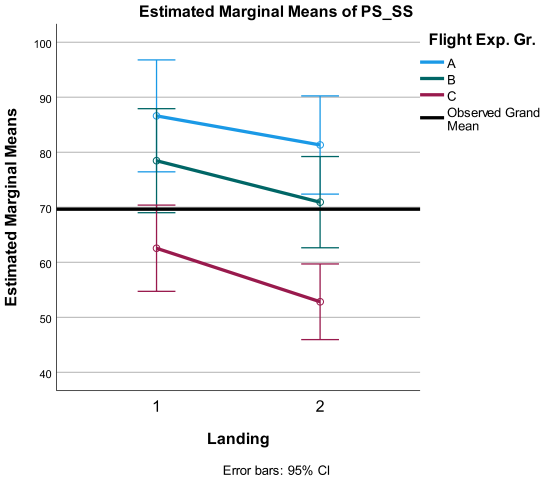


Figure 5.49: Estimated marginal means for PS using SS inceptor, showing differences between two landings (LN and LD scenarios) within each FE group. Landings are coded 1 for LN and 2 for LD.

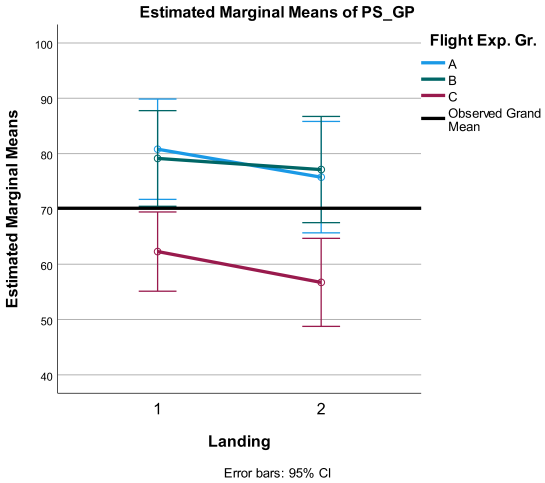


Figure 5.50: Estimated marginal means for PS using GP inceptor, showing differences between two landings (LN and LD scenarios) within each FE group. Landings are coded 1 for LN and 2 for LD.

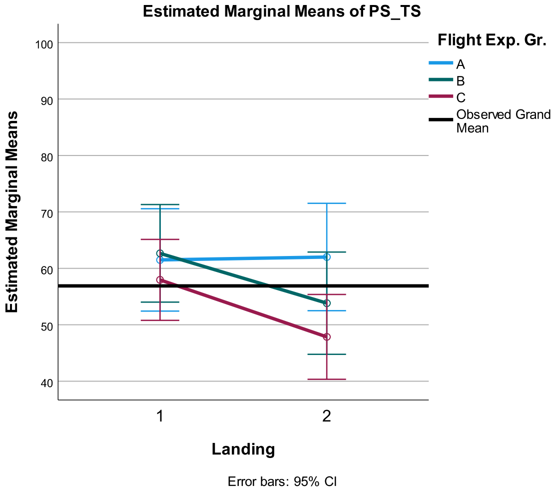


Figure 5.51: Estimated marginal means for PS using TS inceptor, showing differences between two landings (LN and LD scenarios) within each FE group. Landings are coded 1 for LN and 2 for LD.

5.7.15 Summary

This section presents the detailed analyses performed to dissect findings from Section 5.6. In that section, individual grouping factors were nominated as potentially influential on measured variables. Tab. 5.49 presents a visual summary of the findings from this section, along with the effect size of the inceptor. From there, it is evident that inceptors differed significantly in almost all measured factors. Only in perceived supply (SART-S), inceptors' results were similar (the effect was small). This means that participants felt that they were supplied with enough information and understood all three controllers well. This was further confirmed with SUS-L, where, on average, the TS scores were 4 points apart from SS and GP. Moreover, the learnability of all three inceptors was considered at least "good" by the majority of participants[47].

FE was a major factor in some measures - SUS-U, SUS-Total, and all PSs. Even though high-experienced pilots tended to rate the usability of TS as "worst", they managed to complete the scenario tasks using it. This also shows that introducing novel elements in a flight deck might have a big impact on older pilots, while younger pilots would get accustomed quicker to such changes, which is convergent with Taylor & Cotter's research [343]. Interestingly, the fact that the FE group's effect was not significant in CHR, SUS-L, and SART scores means that participants perceived a

---

[47] GP's learnability was rated as "excellent".

similar workload, regardless of their FE, and naïve pilots' SA did not diminish largely from lack of familiarity with a flight deck in general.

It was demonstrated that gender, handedness, and inceptor order did not significantly influence the results, as predicted earlier. Moreover, the results revealed that the attitude towards TS elements in a flight deck, GP and MG usage did not have a significant impact as well.

Playing VG also did not significantly affect the SUS subdimension scores; however, it was found that gamers ranked a GP controller higher. That, along with the previous observation that the SS scores from the pilots were affected by their familiarity with this inceptor [197], are in line with McLellan et al.'s remark that people tend to give higher scores to systems or controllers that they already know [245].

The results proved that randomising the inceptor order ensured no significant learning effect on specific tasks with the given inceptor. On the other hand, having a fixed scenario order provided a slight learning curve. Because of this, the scores in the most challenging LD scenario were not far from those of LN.

All participants performed similarly using a GP in DRH and LN scenarios, proving this inceptor's ergonomic design. The PS was in line with high learnability scores (Section 5.7.3), regardless of the participant's previous experience. The reason for that is the fact that this type of controller has been in development for years (see Section 2.3.7). Moreover, the small difference in performance using SS and GP is convergent with Rupp et al.'s findings [303]. Nonetheless, the usability of GP was rated mostly as a "marginal low" by high-experienced pilots, even though their task performance was similar, and sometimes even better, than while using SS.

Interestingly, the usability scores (SUS-U) of TS were rated much worse by high-experienced pilots compared to low-experienced and naïve pilots, but there was no significant difference between the scores between the latter two groups. The average SUS-U rating from high-experienced pilots was categorised as "worst" usability, while the score from the other two groups could be described as "poor". This proves that older pilots are more reluctant to the introduction of novel research in the cockpit [343]. TS as a controller aside, the younger generation of pilots is more used to TS technology in general, so they might be more interested in seeing such novelties in flight decks. This situation is similar to a mobile phone revolution - nowadays, younger generations are considered "smartphone natives", as smartphones were present in most of their lives, while people who had to learn this technology at an older age are called "smartphone immigrants".

The difference in LN scenario performance was not very large, even though the usability of TS was often rated as "not acceptable". Almost every product must go through a number of iterations before its usability is deemed acceptable, and it can vary during its lifecycle. An example can be found in the article published by Bangor et al., reproduced in Fig. 5.52 [35]. This graph shows that, at the first iteration, the SUS scores tend to be the lowest, only to increase in "upgrade" iterations and then decrease in adding new features or changes, ultimately reaching the highest SUS score after a number of iterations. Because of the unfamiliarity, the TS was also perceived as a controller that was the most difficult to use, according to the CHR scale; on the other hand, NASA-TLX scores suggest that the workload difference among non-gamers was not large, as it only shows an average of 5-point difference on a 100-point scale. The TS controller proposed in this research is only a prototype in its first iteration as an unexplored aircraft inceptor. This creates a possibility for future development. On the other hand, its high learnability shows potential for ground operations in other areas, such as remote/ground control of unmanned aircraft, not necessarily as an "in-flight" aircraft inceptor.

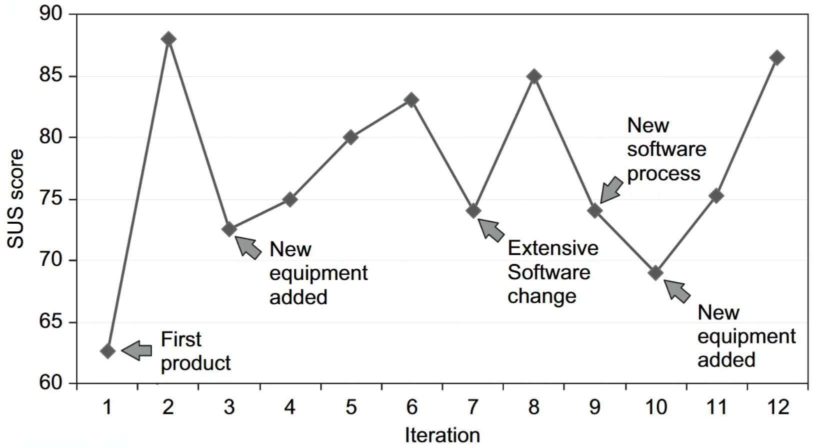


Figure 5.52: SUS scores and their relationship to critical events in the product lifecycle process. Reproduced from Bangor et al.'s article [35].

The analyses from this section allowed validation of the findings from Section 5.6. The results are shown in Tab. 5.49. Effect sizes classification was based on the η_p^2 value obtained from rANOVA results: between 0.01 and 0.06 indicates a small effect; between 0.06 and 0.14 can be interpreted as a medium effect; and 0.14 and higher means a large effect [80].

Table 5.49: Grouping factors after individual analyses, based on Tab. 5.8. Column "Inceptor" was added to indicate the inceptor effect size. The effect sizes are indicated by the letters L (large), Md (medium), and S (small). The significance is marked *** for $p < .001$, ** for $p < .01$, and * for $p < .05$. Effect sizes showing a trend (with elevated tolerance $p < .1$) are marked with $^{\text{T}}$. Rejected or non-significant factors are marked with ×.

Factor	Inceptor	Gender	Handedness	TS attitude	VG freq.	GP usage	MG usage	Inceptor order	FE group	VG group
CHR	L***									
SUS-U	L***			×					L***	×
SUS-L	L***	×					×			×
SUS-Total	L***	×		×					Md$^{\text{T}}$	L$^{\text{T}}$
SART-D	L***					×[48]	×			
SART-S	S*									
SART-U	L***									
SART-Total	L***									
NASA-TLX	L***									Md$^{\text{T}}$
PS DRV	L***	×[49]	×		×[50]		×	×	L***	
PS DRH	L***	×							Md*	
PS LN	L***	×	×			×	×	×	L**	
PS LD	L***	×[49]					×		L***	

Based on the results from this section, another table was created to rank the controllers in each measure scale (Tab. 5.50). The purpose of this table was to provide a clear and easy way to compare the effect size of the inceptor type factor, groups, and interactions and to allow for the drawing of conclusions. The table shows the ranking of the inceptors for each measured factor, including results among significant or trending groups. This allows for an in-depth comparison of the controllers, highlighting their strengths and weaknesses. Additionally, it provides a clear visual representation of the

---

[48] The grouping factor was significant with a medium effect size ($p < .05$), but there were no relevant interactions in GP results.

[49] The grouping factor was significant with a medium effect size ($p < .01$), but the interactions were coincidental.

[50] The grouping factor was significant with a medium effect size ($p < .1$), but there were no relevant interactions.

findings, making it easy to identify patterns and trends. Inceptors were ranked from the best to the worst score. For example, ranking "GP≈SS≫TS" from the CHR measure indicates that participants felt almost the same workload while using SS and GP, while TS induced a significantly higher workload compared to SS and GP. Another example, "GP$\tilde{>}$SS$\tilde{>}$TS", with additional "GP>TS" in the footnote, found in PS (LD) ranking among the participants without flight experience (FE group C) demonstrates that there was a trend of participants performing only slightly better when using GP in comparison to SS, and SS in comparison to TS. The score difference between GP and TS was higher, albeit non-significant.

Table 5.50: Ranking of inceptors for each measured factor, including results among significant or trending groups. Inceptors are ranked from the best to the worst score. The symbols indicate the mean difference (MD) between the inceptors and mean the following: ≫ - significant MD; > - non-significant MD (trend); $\tilde{>}$ - non-significant, small MD; ≈ - insignificant MD. Effect sizes showing a trend (with elevated tolerance $p < .1$) are marked with ^{T}. sc. - scenario; ES - effect size; int. - effect size of interaction between the inceptor and grouping factor; L - large; Md - medium; S - small; ns - not significant.

Measure (sc.)	Factor	Ranking[51]	ES (int.)
CHR	Inceptor	GP≈SS≫TS	L
SUS-U	Inceptor	SS≫GP≫TS	L
	FE gr. A	SS≫GP≫TS	L (Md)
	FE gr. B	SS≫GP≫TS	
	FE gr. C	GP≈SS≫TS	
	VG gr. A	SS$\tilde{>}$GP≫TS	ns (ns)
	VG gr. B	SS≈GP≫TS	
	VG gr. C	SS≫GP≫TS	
SUS-L	Inceptor	GP>SS≫TS	L
	VG gr. A	GP>SS≫TS	ns (ns)
	VG gr. B	GP>SS≫TS	
	VG gr. C	SS>GP>TS[52]	
SUS-Total	Inceptor	SS>GP≫TS	L

continued . . .

---

[52] SS≫TS

Table 5.50: …continued

Measure (sc.)	Factor	Ranking	ES (int.)
SUS-Total	FE gr. A	SS≫GP≫TS	Md$^{\text{T}}$ (Md)
	FE gr. B	SS>GP≫TS	
	FE gr. C	GP>SS≫TS	
	VG gr. A	SS≳GP≫TS	L$^{\text{T}}$ (ns)
	VG gr. B	GP≳SS≫TS	
	VG gr. C	SS≫GP>TS	
SART-D	Inceptor	GP≳SS≫TS	L
SART-S	Inceptor	SS≈GP≈TS	S
SART-U	Inceptor	SS≳GP≫TS	L[53]
SART-Total	Inceptor	SS≈GP≫TS	L[53]
NASA-TLX	Inceptor	GP≫SS≫TS	L
	VG gr. A	GP>SS≫TS	Md$^{\text{T}}$ (ns)
	VG gr. B	GP>SS≫TS	
	VG gr. C	GP≳SS≳TS[54]	
PS (DRV)	Inceptor	GP≫SS≫TS	L
	FE gr. A	GP≫SS≫TS	L (Md)
	FE gr. B	GP≈SS≫TS	
	FE gr. C	GP>SS≫TS	
PS (DRH)	Inceptor	GP≫SS≫TS	L
	FE gr. A	GP≫SS>TS	Md (ns)
	FE gr. B	GP≫SS>TS	
	FE gr. C	GP≫SS≈TS	
PS (LN)	Inceptor	SS≈GP≫TS	L
	FE gr. A	SS>GP≫TS	Md (Md)
	FE gr. B	GP≈SS≫TS	

continued …

---

[53] Large effect, but a small difference in values.

[54] GP>TS

Table 5.50: . . . continued

Measure (sc.)	Factor	Ranking	ES (int.)
PS (LN)	FE gr. C	SS≈GP≳TS[55]	Md (Md)
PS (LD)	Inceptor	GP≈SS≫TS	L
	FE gr. A	SS≳GP≫TS	L (Md)
	FE gr. B	GP≳SS≫TS	
	FE gr. C	GP≳SS≳TS[56]	

---

[55] SS≳TS

[56] GP>TS

6

Summary and conclusions

6.1 Results summary

In order to make the point differences in scores between the inceptors and groups comparable, all scales in this chapter that did not range from 0 to 100[1] were normalised and multiplied by 100. The "%p" symbol used in this section means **percentage points**, not a relative percentage change between two values.

Based on the CHR scale, users experienced approximately 20%p higher workload when using the touchscreen compared to the gamepad and sidestick. Results obtained from NASA-TLX only showed a 12-18%p workload index increase when using the touchscreen, and a 6%p difference between sidestick and gamepad, with the latter causing the least workload during the trials. While both CHR and NASA-TLX are used to measure the workload, NASA-TLX is more detailed and task/human-oriented, as it features six dimensions: mental, physical, and temporal demand, performance, effort, and frustration, while CHR only indicates the participant's effort of completing the task in the given conditions and system (handling qualities). In both scales, results were similar, apart from a small trend in the NASA-TLX scale observed among participants who played video games. Gamers tended to experience the lowest workload when using the gamepad (even among those who had never used this controller before), while non-gamers felt a similar workload among all three inceptors.

The perceived usability of the inceptor relied heavily on the flight experience. Pilots showed a preference for the sidestick by scoring its usability 12-14%p higher than the

---

[1] Those scales were CHR and SART with its subdimensions

gamepad and 40-46%p higher than the touchscreen. On the other hand, non-pilots scored the gamepad and sidestick almost the same, while the touchscreen's usability ratings were 21-26%p lower on average. On the SUS-Total scale, which consists of the usability and learnability measures, the sidestick was ranked only 4%p higher than the gamepad and 30%p higher than the touchscreen. System usability scores can be concluded with a statement that in its current state, the touchscreen is perceived as a controller that is easy to learn but seemingly challenging to use, especially among pilots with significantly more flying experience, while the gamepad offers almost the same experience as the sidestick. Moreover, it was observed that the lower the flight experience was, the higher the gamepad and touchscreen scores tended to get. In an experimental study introducing alternative inceptors in a rotorcraft, Schuchardt observed the same behaviour among professional helicopter pilots. The higher frustration perceived by the pilots was in contrast to their results using alternative controllers, which was explained by their "conservativeness" caused by years of experience [317].

Learnability analyses showed that, while there was only a 5%p difference between the gamepad and sidestick, the first one was ranked as an inceptor with "excellent" learnability, while the second fell into a "good" learnability category. Interestingly, the touchscreen was also categorised as having "good" learnability, even though its learnability scores were 17-22%p lower than the other two inceptors. Furthermore, a trend was observed that gamers gave higher learnability scores for a gamepad.

In the demand subdimension of SART, participants felt that the touchscreen was 14%p more demanding to use compared to the sidestick and gamepad, while the other two were similar (with the latter seemingly more demanding, though). Perceived supply (SART-S) and understanding were similar among all three inceptors, with scores differing by just 4%p for the SART-S and 11%p for the SART-U. This indicates that the participants felt well-informed and clearly understood the controllers (which supports their "good" rating in the learnability measure). The increased demand among all three inceptors was counterbalanced by perceived supply and understanding. Overall, the situation awareness (SA), calculated by subtracting the demand from the supply and understanding scores, was moderate and similar among all three inceptors, with the ratings of the sidestick and gamepad being only 5%p higher than those of the touchscreen. Moreover, the grouping factors did not significantly affect the SART ratings. This is interesting, as higher SA is usually associated with pilot training, experience, and abilities [116, 381].

Contrary to the participants' subjective ratings, the gamepad's objective performance was the highest in the disturbance rejection scenarios, even among experienced pilots. In the DRV[2] scenario, the average Performance Scores (PS) acquired with the gamepad were 7%p higher than with the sidestick. Surprisingly, the score difference between the gamepad and sidestick was the highest among the high-experienced pilots' group (12%p). Touchscreen results were significantly lower, but the PS between pilots and non-pilots using this controller was only 8%p-14%p higher for high- and low-experienced pilots, respectively (as compared to non-pilots). Interestingly, pilots with lower experience achieved 6%p better scores using the touchscreen compared to their more experienced colleagues, but the gamepad scores were higher by 9%p for the latter, even though there were fewer gamepad users among that group. DRH[3] scenario was seen as the easiest to complete, and the results confirmed this: average PS ranged from the touchscreen's $PS = 71$, through the sidestick's $PS = 76$, up to the gamepad's $PS = 88$. Interestingly, non-pilots' performance with the sidestick was only 1%p higher than with the touchscreen, and the gamepad results were 17-18%p higher than the other two inceptors.

Results from landing scenarios showed a similarity in performance achieved with the sidestick and gamepad. Overall, the sidestick's PS in LN[4] scenario was only 1%p higher than the gamepad's, and both inceptors were 12-13%p higher than the touchscreen. Participants without flight experience achieved similar results with every inceptor - the average scores ranged from $PS = 58$ (touchscreen) to $PS = 63$ (sidestick and gamepad). Moreover, the results achieved with touchscreen, regardless of piloting experience, ranged from $PS = 58$ to $PS = 63$[5]. In LD[6] task, the patterns were similar to those in the LN scenario (gamepad's PS was 2%p higher than sidestick's, and touchscreen results were lower by 13-15%p), but the influence of flight experience was more apparent: high-experienced pilots' scores were 11%p higher than participants' with low flight experience, and 29%p higher than non-pilots'. The range of PS scores among non-pilots was similar, ranging from $PS = 48$ (touchscreen) to $PS = 57$ (gamepad), with sidestick in the middle. The investigation of touchscreen results showed that, in this case, flight experience also mattered. High-experienced pilots achieved the best scores compared to low-experienced and naïve pilots (by 8-14%p, respectively). Per-

---

[2] Disturbance rejection in the vertical channel.

[3] Disturbance rejection in the horizontal channel.

[4] Landing with no disturbance.

[5] Coincidentally, the range was the same as for naïve pilots with any inceptor.

[6] Landing with disturbance.

haps the most interesting observation is that high-experienced pilots achieved the same score in both landing scenarios using the touchscreen, despite the significantly greater difficulty of the second task. This proves that pilot training and substantial flying experience help pilots quickly adapt to novel prototype inceptors.

6.2 Conclusions

This summary provides valuable insights into the potential of developing alternative inceptors for aircraft. The gamepad, in particular, showed promising results in terms of performance, even though participants rated its usability lower than that of the sidestick. This suggests that ergonomic design, while important, is not the only factor that affects the user's experience and the system's performance, as device familiarity and occupational training are also significant aspects. On the other hand, the touchscreen did not perform as well as the other two controllers and, being still in its infancy as a flight deck interface, is not yet viable to be considered an alternative inceptor. However, it received good scores in learnability and understanding scales, which suggests that it may have potential as an off-board controller in other areas than a flight deck, such as unmanned aircraft. Interestingly, participants with no flight experience performed similarly using all three inceptors, which suggests that, with the proper training, both proposed **alternative inceptors can potentially improve the pilot's performance**, supporting the hypothesis (H_1) defined in Section 1.4.

Further to the alternative inceptors and to appreciate the importance of human factors in aviation, the research conducted in this thesis aimed to investigate the effect of different participant characteristics in the flight simulator. The analyses revealed that flight experience had the most significant impact on the participants' behaviour. Additionally, a trend was observed that video game experience might influence performance, particularly among non-pilots, and also impacts the perceived usability of the gamepad controller, proving the matter of a ”familiarity bias” exists. Moreover, it was demonstrated that it is worth considering the utilisation of the user-centred design when developing a novel engineering flight simulator. These findings support the hypothesis (H_3) that **demographic, occupational, and personal characteristics have a significant effect on the subjective experience and objective performance in the flight simulator**.

The development of a state-of-the-art engineering flight simulator called the Future

Systems Simulator (FSS) was a crucial element in the research and validation of alternative inceptors for aircraft control. The simulator's flexibility and modularity allowed for conducting experiments in a range of simulated scenarios. Furthermore, the fact that the FSS's interface was coded entirely by the author allowed the simulator to be highly customisable and enabled full control over the trial procedures. The results obtained through the use of the FSS supplied valuable insights into the effectiveness of alternative inceptors, as well as the impact of human factors on the results, supporting the investigation of hypotheses (H_1) and (H_3). The development of the FSS provided a powerful tool to design and carry out the experiments and obtain credible and satisfactory results, thus confirming the hypothesis (H_2) that **a novel engineering flight simulator helps to streamline the research and validate the results of radically different control methods in an aircraft**. This hypothesis was also further supported by other research activities performed in the FSS [103, 137, 198, 216–219], addressed in Sections 3.5 and 6.4.

6.3 Limitations and recommendations for further research

The study focused on the physical design of alternative inceptors and the pilot interface in an engineering flight simulator. Therefore, the aircraft dynamics or stability and control methods were out of scope. Moreover, alternative rudder control was not considered, as it is not used often by pilots. Those matters were addressed in Sections 4.4 and 4.5. The FSS has a fixed base and does not include a motion system, so the disturbance was only simulated in the form of a digital signal injected into the inceptor's input. This was described in more detail in Section 3.4.5.

None of the controller types was an active inceptor, and therefore no tactile feedback was provided to the pilots. The significant difference, however, was in the sense of control - with sidestick and gamepad, the pilot "knew" if they had reached the limit of the displacement distance without looking down at the controller (because the controller was physically limited by design). With the touchscreen controller, no physical border indicated the limits. This issue could be addressed in future trials by adding a small bulge around the TS controller area, for example, by using a 3D printer.

Due to the software limitation, multi-touch was not supported on the experimental rig, so only single-tapping and sliding gestures were used, as described in Section 4.4.3. However, in future experiments, it is recommended to perform the trials using multi-touch displays and full-motion platforms in order to assess the pilot's performance in high-turbulence situations.

This particular research did not need to assure the accuracy of the touch – the proposed touch controller is not in close contact with other touchscreen elements. Hence, initiating an interaction did not require high accuracy, only constant holding of the control "knob". This introduced fatigue in the arm and wrists of some participants. Fatigue is an acknowledged issue in human-computer interaction research [325, 331, 377].

It was determined that the touchscreen controller is not yet ready to be used as an onboard inceptor. Nonetheless, it was found that touchscreens offer many advantages that could be beneficial in other fields of aviation. Moreover, using a gamepad as a controller proved to offer a similar performance to the sidestick, even among professional pilots and despite the fact that some of them were not familiar with this controller before. This section presents some examples of further research based on the outcomes of this research and participants' comments.

To better understand the pilot's behaviour and the inceptor's performance, additional off-the-shelf devices could be implemented, such as eye-tracking [198], heart rate variability (HRV) monitor [222] or electroencephalography (EEG) [29]. These solutions are meaningful when assessing the pilot's workload – where the pilot is looking at a specific point of time during the simulation or which situations cause the highest stress rates. In this study, some of the eye-tracking and HRV data was collected, but not for all participants; hence it was not included in the main analyses. Filtering participants' results to investigate their gaze patterns and stress levels will be done in the future.

One of the propositions to further examine the touchscreen as an inceptor is a use of "bendable" touchscreens, allowing to create an inward-bent conical panel on the sidestick pedestal, where the centre of the cone would be a central controller point. Additional cavities or recesses, running along the X and Y axes, could give a pilot tactile feedback on where his finger would be at the moment. Folding screen ideas and prototypes were already present as early as 2007, albeit they were not very successful due to the technological limitations [16]. Nowadays, mobile phones like Galaxy Fold and Flip [310] with foldable screens are widely commercially available. This allows to speculate that in the foreseeable future, it will be possible to shape a touch display in any way. In the case of touchscreen controllers, it will enable the implementation of "haptic" cone-shaped controller, in which the centre can be the lowest part, indicating the idle position and satisfying "context-aware" and "eyes-free" properties, proposed by Castillo & Couture [62], as well as "Function Equals Form" theme in Organic User Interfaces design, defined by Vertegaal & Poupyrev [368]. Moreover, further trials might include a touchscreen inceptor implemented on Galaxy Flip (or similar) smartphone, placed on a sidestick pedestal. This would ensure a familiar positioning for the pilots and would not obstruct the primary flight display. The folding feature would allow for more tactile control, similar to the "accordion" design proposed by Pauchet et al. [280].

Touchscreen controller operation can be further enhanced with Machine Learning technology, making it adaptive according to the user's actions, similar to the Smart controller proposed by Torok et al. [358]. A prototype designed for AI-supported tangible car control, proposed by Ghani et al. [135], could be enhanced with an addition of a touchscreen to the single moving elements and adapted for an aircraft controller in order to create a tactile feeling to the pilot.

Alternatively to using a touchscreen as a primary controller, it could be investigated as a visual aid in the adjustment of the flight path. Due to the fact that the touchscreen can offer intuitive input and immediate response, it could be used as a supplemental system in autopilot procedures. For example, a pilot could adjust the flight path shown on the navigation display by simply sliding a finger towards a desired position.

Although the results showed that flight experience had the largest impact on pilot performance and usability ratings, there was still a tendency suggesting that video game experience might play a role in training pilots using alternative inceptors. The results indicated that gamers tend to learn unfamiliar controllers more quickly than non-gamers. It is therefore recommended to consider the role of video game experience as a potential factor in pilot training when using alternative inceptors, especially in the context of new technologies such as touchscreen controllers. Due to the minor role of the video game factor in this study, the information about the video game genres played by participants was not analysed in this study. However, in future, it is worth considering this, as, for example, Dobrowolski et al. recognised the importance of different genres in video game studies and showed that playing real-time strategy (RTS) games increases performance on multiple object tracking, and that there are behavioural differences between people playing specific game genres (first-person shooters versus RTS) [105].

Both the gamepad and touchscreen can be studied as inceptors for urban air mobility vehicles. There is evidence indicating the need for specialised inceptors for unmanned aircraft. Research has not looked into touchscreen options, focusing on passive and active sticks, mouse-like devices, and gamepads [129, 306]. The gamepad and touchscreen controllers proposed in this study merit further examination as potential alternatives for unmanned aircraft inceptor systems.

The development of the FSS showed that building a new engineering flight simulator is a very complex process and it requires many considerations in the early design, from hardware to human factors aspects. Therefore, it is important to think forward when defining requirements for the flight simulator. Hardware and network architecture should be considered as early as possible, including aspects such as the physical locations of the PCs and the cable types and lengths. Moreover, from the software design perspective, a style guide and, if Unity is used, "prefabs" for the user interface elements should be designed and utilised throughout the development to keep the consistency of the system.

6.4 Published and submitted materials

Journal articles

- Li, W.-C., **Korek, W. T.**, Liang, Y. H., & Lin, J. J. H. (2023). "Touchscreen Controls for Future Flight Deck Design: Investigating Visual Parameters on Human-Computer Interactions between Pilot Flying and Pilot Monitoring". *Journal of Aeronautics, Astronautics and Aviation*, Vol. 55(2), 201–211. ISSN: 2352-1465. DOI: 10.6125/JoAAA.202306_55(2).08

- Li, W.-C., Wang, Y., & **Korek, W. T.** (2022). "To be or not to be? Assessment on using touchscreen as inceptor in flight operation". *Transportation Research Procedia*, Vol. 66, 117–124. DOI: 10.1016/j.trpro.2022.12.013

Conference papers

- **Korek, W. T.**, Li, W.-C., Lu, L., & Lone, M. (2022). "Investigating Pilots' Operational Behaviours While Interacting with Different Types of Inceptors". In D. Harris & W. Li (Eds.), *Lecture Notes in Computer Science (including subseries*

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 13307 LNAI, 314–325. Springer, Cham. ISBN: 978-3-031-06086-1. DOI: 10.1007/978-3-031-06086-1_24. International Conference on Human-Computer Interaction (HCII 2022), 26.06-1.07.2022 (online).

- Li, W.-C., Liang, Y. H., **Korek, W. T.**, & Lin, J. J. H. (2022). ”Assessments on Human-Computer Interaction Using Touchscreen as Control Inputs in Flight Operations”. In D. Harris & W. Li (Eds.), *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 13307 LNAI, 326–338. Springer, Cham. ISBN: 978-3-031-06086-1. DOI: 10.1007/978-3-031-06086-1_25. International Conference on Human-Computer Interaction (HCII 2022), 26.06-1.07.2022 (online).

- Li, W.-C., Liang, Y., & **Korek, W. T.** (2022). ”Flight operations using touchscreen controls: assessing system usability and pilots' visual attention”. In N. Balfe & D. Golightly (Eds.), *Contemporary Ergonomics & Human Factors 2022: Proceedings for the Annual Conference of the Chartered Institute of Ergonomics and Human Factors.* CIEHF. ISBN: 978-1-9996527-4-6. Ergonomics & Human Factors International Conference (EHF 2022), 11-12.04.2022 (online); 25-26.04.2022 (Birmingham, UK).

- **Korek, W. T.**, Mendez, A., Asad, H. U., Li, W.-C., & Lone, M. M. (2020). ”Understanding Human Behaviour in Flight Operation Using Eye-Tracking Technology”. In D. Harris & W. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics. Cognition and Design. HCII 2020. Lecture Notes in Computer Science*, Vol. 12187, 304–320. Springer, Cham. ISBN: 978-3-030-49183-3. DOI: 10.1007/978-3-030-49183-3_24. International Conference on Human-Computer Interaction (HCII 2020), 19-24.07.2020 (online).

Manuscripts accepted for publication or submitted for revision

- Li, W.-C., Wang, Y., **Korek, W. T.**, & Braithwaite, G. (2023). ”Future flight deck design: Implementation of a touchscreen as flight inceptor for single pilot operation”. *International Journal of Human-Computer Studies.* Manuscript under revision, awaiting preprint DOI. Dataset available at 10.17862/cranfield.rd.21907797

- Hu, K., Li, W.-C., & **Korek, W. T.** (2023). "Assessing pilots' situation awareness using touchscreen inceptor". Manuscript accepted, to be presented at the Ergonomics & Human Factors International Conference (EHF 2022), 24-26.04.2023 (Kenilworth, UK). Proceedings will be published as *Contemporary Ergonomics & Human Factors 2022: Proceedings for the Annual Conference of the Chartered Institute of Ergonomics & Human Factors*. Publisher: CIEHF.

Bibliography

[1] Abraham, D., Fergus, M., and Franck, T., *The Expanse*, [TV series]. USA: Syfy (s. 1-3); Amazon Prime Video (s. 4-6), 2015.

[2] Abzug, M. J. and Larrabee, E. E., *Airplane Stability and Control, Second Edition*. Cambridge: Cambridge University Press, 2002, ISBN: 9780511607141. DOI: `10.1017/CBO9780511607141`.

[3] Administration Federal Aviation, "Airplane simulator qualification," Tech. Rep., 1991, pp. 5–12.

[4] Affonso, W., Tavares, R., Barbosa, F. R., Gandolfi, R., Reis, R. J. N. dos, Silva, C. R. I. da, Kipouros, T., Laskaridis, P., Enalou, H. B., Chekin, A., Kukovinets, A., Gubernatorov, K., Ravikovich, Y., Ivanov, N., Ponyaev, L., and Holobtsev, D., "System architectures for thermal management of hybrid-electric aircraft - FutPrInt50," *IOP Conference Series: Materials Science and Engineering*, vol. 1226, no. 1, p. 012 062, 2022, ISSN: 1757-8981. DOI: `10.1088/1757-899X/1226/1/012062`.

[5] Airbus, *Airbus brings cockpit to you with new Virtual Reality Flight Trainer*, 2019. [Online]. Available: `https://aircraft.airbus.com/en/airbus-brings-cockpit-to-you-with-new-virtual-reality-flight-trainer` (visited on 09/16/2022).

[6] Alankuş, G. and Eren, A. A., "Enhancing gamepad fps controls with tilt-driven sensitivity adjustment," *Contemporary Topics in Computer Graphics and Games*, pp. 211–230, 2019.

[7] Alapetite, A., Møllenbach, E., Stockmarr, A., and Minakata, K., "A Rollercoaster to Model Touch Interactions during Turbulence," *Advances in Human-Computer Interaction*, vol. 2018, pp. 1–16, 2018, ISSN: 1687-5893. DOI: `10.1155/2018/2698635`.

[8] Alba-Maestre, J., Prud'homme van Reine, K., Sinnige, T., and Castro, S. G. P., "Preliminary Propulsion and Power System Design of a Tandem-Wing Long-Range eVTOL Aircraft," *Applied Sciences*, vol. 11, no. 23, p. 11 083, 2021, ISSN: 2076-3417. DOI: `10.3390/app112311083`.

[9] Alexander, M. and Chen, H.-l., "Perceptions and effects of a system's usability by experience level," *Proceedings of the American Society for Information Science and Technology*, vol. 40, no. 1, pp. 389–397, 2005, ISSN: 00447870. DOI: `10.1002/meet.1450400147`.

[10] Allerton, D. J., "Flight simulation - Past, present and future," *Aeronautical Journal*, vol. 104, no. 1042, pp. 651–663, 2000, ISSN: 00019240. DOI: `10.1017/s0001924000096901`.

[11] Allerton, D. J., "The impact of flight simulation in aerospace," *Aeronautical Journal*, vol. 114, no. 1162, pp. 747–756, 2010, ISSN: 00019240. DOI: `10.1017/S0001924000004231`.

[12] Allerton, D., *Principles of Flight Simulation.* Wiley, 2009, pp. 1–471, ISBN: 9780470754368. DOI: `10.1002/9780470685662`.

[13] Andonian, B., Rauch, W., and Bhise, V., "Driver steering performance using Joystick vs. Steering wheel controls," in *SAE Technical Papers*, 2003. DOI: `10.4271/2003-01-0118`.

[14] Anon, "The Terra Tutor," *Aviation and Aeronautical Engineering*, vol. 3, no. 8, p. 532, 1917.

[15] Anon, "Military Spefication Flying Qualities of Piloted Airplanes, MIL-STD-1797A," Department of Defense, USA, Tech. Rep., 1969.

[16] Anon, *Polymer Vision Readius*, 2014. [Online]. Available: `https://www.mobilegazette.com/polymer-vision-readius-07x02x10.htm` (visited on 11/20/2022).

[17] Anon, *Garmin Receives EASA Approval of the G3X Touch Flight Display in Single-Engine Piston Aircraft*, 2020. [Online]. Available: `https://www.aviationpros.com/engines-components/aircraft-airframe-accessories/avionics/press-release/21146910/garmin-international-garmin-receives-easa-approval-of-the-g3x-touch-flight-display-in-singleengine-piston-aircraft` (visited on 12/21/2022).

[18] Anon, *Flight Simulators of Yesteryear*, 2022. [Online]. Available: `https://www.aerotoons.com/blog/yesteryear%5C_sims.html` (visited on 09/15/2022).

[19] Antoinette, *Training rig for Antoinette aircraft (historic photo)*, 1909. [Online]. Available: `https://en.wikipedia.org/wiki/File:Antoinette%5C_sim%5C_2.jpg` (visited on 09/10/2022).

[20] Apple, *Apple Reinvents the Phone with iPhone*, 2007. [Online]. Available: `https://www.apple.com/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/` (visited on 09/22/2022).

[21] Arab, F., Malik, Y., and Abdulrazak, B., "Evaluation of PhonAge: An Adapted Smartphone Interface for Elderly People," in *Human-Computer Interaction – INTERACT 2013. Lecture Notes in Computer Science*, Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., and Winckler, M., Eds., vol. 8120, 2013, pp. 547–554. DOI: `10.1007/978-3-642-40498-6_44`.

[22] Arthur III, J., Shelton, K., Prinzel III, L., and Bailey, R., "Performance Evaluation of Speech Recognition Systems as a Next-generation Pilot-vehicle Interface Technology," Tech. Rep., 2016.

[23] Asmayawati, S. and Nixon, J., "Modelling and supporting flight crew decision-making during aircraft engine malfunctions: developing design recommendations from cognitive work analysis," *Applied Ergonomics*, vol. 82, p. 102 953, 2020, ISSN: 00036870. DOI: `10.1016/j.apergo.2019.102953`.

[24] Astin, A. and Nussbaum, M. A., "Interactive Effects of Physical and Mental Workload on Subjective Workload Assessment," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 13, pp. 1100–1104, 2002, ISSN: 2169-5067. DOI: `10.1177/154193120204601320`.

[25] Avsar, H., Fischer, J., and Rodden, T., "Target size guidelines for interactive displays on the flight deck," in *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, IEEE, 2015, pp. 3C4–1–3C4–15, ISBN: 978-1-4799-8940-9. DOI: `10.1109/DASC.2015.7311400`.

[26] Baarspul, M., "A review of flight simulation techniques," *Progress in Aerospace Sciences*, vol. 27, no. 1, pp. 1–120, 1990, ISSN: 03760421. DOI: `10.1016/0376-0421(90)90006-6`.

[27] Bachelder, E. N. and Aponso, B. L., "Application of the Perceptual-Behavioral Pilot Model for Handling Qualities Assessment and Prediction," in *AIAA SCITECH 2022 Forum*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2022, ISBN: 978-1-62410-631-6. DOI: `10.2514/6.2022-0890`.

[28] Baguley, R., *The Gadget We Miss: The Nokia 9000 Communicator*, 2013. [Online]. Available: `https://medium.com/people-gadgets/the-gadget-we-miss-the-nokia-9000-communicator-ef8e8c7047ae` (visited on 09/22/2022).

[29] Bai, Y., Liu, T., Zhang, J., and Luo, Y., "Design and Implementation of a Flight Simulation Platform for EEG Experiment," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, IEEE, 2017, pp. 426–429, ISBN: 978-1-5386-3021-1. DOI: `10.1109/IHMSC.2017.104`.

[30] Bajka, M., Tuchschmid, S., Fink, D., Székely, G., and Harders, M., "Establishing construct validity of a virtual-reality training simulator for hysteroscopy via a multimetric scoring system," *Surgical Endoscopy*, vol. 24, no. 1, pp. 79–88, 2010, ISSN: 0930-2794. DOI: `10.1007/s00464-009-0582-4`.

[31] Baldauf, M., Fröhlich, P., Adegeye, F., and Suette, S., "Investigating On-Screen Gamepad Designs for Smartphone-Controlled Video Games," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 1s, pp. 1–21, 2015, ISSN: 1551-6857. DOI: `10.1145/2808202`.

[32] Baldus, T. and Patterson, P., "Usability of pointing devices for office applications in a moving off-road environment," *Applied Ergonomics*, vol. 39, no. 6, pp. 671–677, 2008, ISSN: 00036870. DOI: `10.1016/j.apergo.2008.01.004`.

[33] Balouchi, F., *Winners of the IDTechEx future of electric vehicles awards*, 2010. [Online]. Available: `https://www.idtechex.com/fr/research-article/winners-of-the-idtechex-future-of-electric-vehicles-awards/2924` (visited on 09/20/2022).

[34] Bandyopadhyay, A., Raj, N. S. S., and Varghese, J. T., "Coexisting in a world with urban air mobility: A revolutionary transportation system," in *2018 Advances in Science and Engineering Technology International Conferences (ASET)*, IEEE, 2018, pp. 1–6, ISBN: 978-1-5386-2399-2. DOI: `10.1109/ICASET.2018.8376817`.

[35] Bangor, A., Kortum, P. T., and Miller, J. T., "An Empirical Evaluation of the System Usability Scale," *International Journal of Human-Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008, ISSN: 1044-7318. DOI: `10.1080/10447310802205776`.

[36] Bangor, A., Kortum, P. T., and Miller, J. T., "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale - International Journal of Usability Studies," *Journal of Usability Studies*, vol. 4, no. 3, pp. 113–123, 2009. DOI: `10.5555/2835587.2835589`.

[37] Bardi, J. S., *The Calculus Wars : Newton, Leibniz, and the Greatest Mathematical Clash of All Time.* New York: Thunder's Mouth Press, 2006, ISBN: 1-56025-706-7.

[38] Basket, B. J., "ADS-33E-PRF Aeronautical Design Standard Performance Specification Handling Qualities Requirements for Military Rotorcraft," Aviation Engineering Directorate, Alabama, Tech. Rep., 1996.

[39] Bauer, M. and Klingauf, U., "Virtual-Reality as a Future Training Medium for Civilian Flight Procedure Training," in *AIAA Modeling and Simulation Technologies Conference and Exhibit*, Reston, Virigina: American Institute of Aeronautics and Astronautics, 2008, pp. 1–7, ISBN: 978-1-62410-000-0. DOI: `10.2514/6.2008-7030`.

[40] Baxter, G., Besnard, D., and Riley, D., "Cognitive mismatches in the cockpit: Will they ever be a thing of the past?" *Applied Ergonomics*, vol. 38, no. 4, pp. 417–423, 2007, ISSN: 00036870. DOI: `10.1016/j.apergo.2007.01.005`.

[41] Bécouarn, L., Dominici, J., Bader, J., Fabbri, M., Pregnolato, M., Sarayeddine, K., Cuypers, D., De Smet, H., Alapetite, A., Sgouros, N., Kouros, P., Zammit-Mangion, D., and Pace, M. F., "One DIsplay for a Cockpit Interactive Solution (ODICIS) Project final report," Thales Avionics, Tech. Rep., 2012.

[42] Beeks, D., "Speech recognition and synthesis," in *Digital Avionics Handbook*, Spitzer, C., Ferrell, U., and Ferrell, T., Eds., 3rd ed., CRC Press, 2014.

[43] Billings, C. E., *Aviation Automation: The Search for A Human-centered Approach.* CRC Press, 2018.

[44] Blake, M. W., "The NASA advanced concepts flight simulator: A unique transport aircraft research environment," *1996 Flight Simulation Technologies Conference*, pp. 385–392, 1996. DOI: `10.2514/6.1996-3518`.

[45] Blake, X. J., Hongdilokkul, R., and Aphiratsakun, N., "Halocode-based Gamepad Development," in *2022 7th International STEM Education Conference (iSTEM-Ed)*, IEEE, 2022, pp. 1–4, ISBN: 978-1-6654-9821-0. DOI: `10.1109/iSTEM-Ed55321.2022.9920764`.

[46] Blundell, J., Huddlestone, J., Collins, C., Scott, S., Sears, R., and Plioutsias, A., "Envisioning Mixed Realities on the Flight Deck," in 2020, pp. 493–499. DOI: `10.1007/978-3-030-50943-9_62`.

[47] Boeing, *Flying Virtually Solo*, 2017. [Online]. Available: `https://www.boeing.com/features/2017/07/virtual-copilot-07-17.page` (visited on 09/16/2022).

[48] Bolton, M., "Blowing in the Wind: Eardley Billing's Oscillator and its Successors," *Journal of Aeronautical History*, no. 06, pp. 166–199, 2018.

[49] Bonaiuto, S., Cannavo, A., Piumatti, G., Paravati, G., and Lamberti, F., "Teleoperation of Robot Teams: A Comparison of Gamepad-, Mobile Device and Hand Tracking-Based User Interfaces," in *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, IEEE, 2017, pp. 555–560, ISBN: 978-1-5386-0367-3. DOI: `10.1109/COMPSAC.2017.278`.

[50] Borkowski, M., Modzelewska, R., Siluszyk, M., Iskra, K., and Zienkiewicz, T., "Possible relationship between solar activity and disturbances in the communication and radar operation on the air traffic control tower of the military airport in Deblin (Poland)," in *43rd COSPAR Scientific Assembly. Held 28 January - 4 February*, vol. 43, 2021, p. 645.

[51] Boslaugh, S. and Watters, P. A., *Statistics in a Nutshell: A Desktop Quick Reference*, Treseler, M., Ed. Sebastopol, CA: O'Reilly Media, Inc., 2008, ISBN: 978-0-596-51049-7.

[52] Bouchner, P., "Interactive Driving Simulators - History, Design and their Utilization in area of HMI Research," *International journal of systems applications, engineering & developmant*, vol. 10, pp. 179–188, 2016, ISSN: 1388-6150.

[53] Bréda, R. and Adamčík, F., "Aircraft automatic control systems and their control systems," *Naše more*, vol. 61, no. 1-2, pp. 9–12, 2014, ISSN: 04696255.

[54] Brooke, J., "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation In Industry*, November 1995, CRC Press, 1996, pp. 207–212. DOI: `10.1201/9781498710411-35`.

[55] Bulusu, V. and Sengupta, R., "Urban Air Mobility: Viability of Hub-Door and Door-Door Movement by Air," 2020. DOI: `https://doi.org/10.7922/G2QJ7FK0`.

[56] Burgess-Limerick, R., Zupanc, C. M., and Wallis, G., "Directional control–response compatibility of joystick steered shuttle cars," *Ergonomics*, vol. 55, no. 10, pp. 1278–1283, 2012, ISSN: 0014-0139. DOI: `10.1080/00140139.2012.700328`.

[57] Bustamante, E. A. and Spain, R. D., "Measurement Invariance of the Nasa TLX," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 52, no. 19, pp. 1522–1526, 2008, ISSN: 2169-5067. DOI: `10.1177/154193120805201946`.

[58] Cantu, A., Vinot, J. L., Letondal, C., Pauchet, S., and Causse, M., "Does folding improve the usability of interactive surfaces in future airliner cockpits? An evaluation under turbulent conditions and varying cognitive load," in *32ème Conférence Internationale Francophone sur l'Interaction Homme-Machine, IHM 2021 - Actes de la Conference*, 2021, pp. 1–10, ISBN: 9781450383622. DOI: `10.1145/3450522.3451246`.

[59] Cao, Y., "Design and Implementation of Certain Type Flight Test Simulation Platform Visual System," in *Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control*, New York, NY, USA: ACM, 2020, pp. 1–5, ISBN: 9781450388894. DOI: `10.1145/3440084.3441203`.

[60] Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., and Göbel, S., "Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook," *Virtual Reality*, vol. 25, no. 4, pp. 1153–1170, 2021, ISSN: 1359-4338. DOI: `10.1007/s10055-021-00513-6`.

[61] Casiez, G., Vogel, D., Balakrishnan, R., and Cockburn, A., "The Impact of Control-Display Gain on User Performance in Pointing Tasks," *Human-Computer Interaction*, vol. 23, no. 3, pp. 215–250, 2008, ISSN: 0737-0024. DOI: `10.1080/07370020802278163`.

[62] Castillo, J. A. L. del and Couture, N., "The aircraft of the future," in *Proceedings of the International Conference on Human-Computer Interaction in Aerospace*, New York, NY, USA: ACM, 2016, pp. 1–8, ISBN: 9781450344067. DOI: `10.1145/2950112.2964582`.

[63] Chai, X., Yu, X., and Wang, Y., "Tradeoff Study between Cost and Environmental Impact of Aircraft Using Simultaneous Optimization of Airframe and Engine Cycle," *International Journal of Aerospace Engineering*, vol. 2017, pp. 1–10, 2017, ISSN: 1687-5966. DOI: `10.1155/2017/2468535`.

[64] Chan, W. T.-K. and Li, W.-C., “Consideration of Cultural Differences in Future Workplace Design for Single Pilot Operations,” in *CIEHF Ergonomics & Human Factors 2021*, 2021, pp. 1–2.

[65] Chan, W. T.-K. and Li, W.-C., “Cultural Effects on the Selection of Aviation Safety Management Strategies,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13307 LNAI, 2022, pp. 245–252, ISBN: 9783031060854. DOI: `10.1007/978-3-031-06086-1_18`.

[66] Chan, W. T.-K. and Li, W.-C., “Investigating professional values among pilots, cabin crew, ground staff, and managers to develop aviation safety management systems,” *International Journal of Industrial Ergonomics*, vol. 92, p. 103 370, 2022, ISSN: 01698141. DOI: `10.1016/j.ergon.2022.103370`.

[67] Changeon, G., Graeff, D., Anastassova, M., and Lozada, J., “Tactile emotions: A vibrotactile tactile gamepad for transmitting emotional messages to children with autism,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, PART 1, vol. 7282 LNCS, 2012, pp. 79–90. DOI: `10.1007/978-3-642-31401-8_8`.

[68] Chappell, S. L. and Sexton, G. A., “Advanced concepts flight simulation facility,” *Applied Ergonomics*, vol. 17, no. 4, pp. 252–256, 1986, ISSN: 00036870. DOI: `10.1016/0003-6870(86)90126-2`.

[69] Chen, C. X. and Mao, Y. H., “Design of Remote Control Aircraft Model Based on ARM,” *Applied Mechanics and Materials*, vol. 568-570, pp. 954–957, 2014, ISSN: 1662-7482. DOI: `10.4028/www.scientific.net/AMM.568-570.954`.

[70] Chen, H.-J., Lin, C. J., and Lin, P.-H., “Effects of control-display gain and postural control method on distal pointing performance,” *International Journal of Industrial Ergonomics*, vol. 72, pp. 45–53, 2019, ISSN: 01698141. DOI: `10.1016/j.ergon.2019.04.004`.

[71] Chu, J., Wu, L., and Xu, J., *Touch Screen Application in Flight Deck for Civil Aircraft*. Springer International Publishing, 2021, vol. 275, pp. 667–675, ISBN: 9783030800901. DOI: `10.1007/978-3-030-80091-8_79`.

[72] Churn, P., Maxwell, C., Schofield, N., Howe, D., and Powell, D., “Electrohydraulic actuation of primary flight control surfaces,” in *IEE Colloquium on All Electric Aircraft*, vol. 1998, IEE, 1998, pp. 3–3. DOI: `10.1049/ic:19980341`.

[73] Cieslak, J., Zolghadri, A., Goupil, P., and Dayre, R., "A Comparative Study of Three Differentiation Schemes for the Detection of Runaway Faults in Aircraft Control Surfaces," *IFAC-PapersOnLine*, vol. 49, no. 17, pp. 70–75, 2016, ISSN: 24058963. DOI: `10.1016/j.ifacol.2016.09.013`.

[74] Clarke, M., Smart, J., Botero, E. M., Maier, W., and Alonso, J. J., "Strategies for Posing a Well-Defined Problem for Urban Air Mobility Vehicles," in *AIAA Scitech 2019 Forum*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2019, ISBN: 978-1-62410-578-4. DOI: `10.2514/6.2019-0818`.

[75] Clarkson, J., Writer, and Whitehead, G., Producer, *Clarkson's Farm (Season 1, Episode 1): "Tractoring"*, Fincham, P. and Wilman, A., Executive producers, [TV series episode]. UK: Amazon Prime Video, 2021.

[76] Clarkson, R. M., "High-Speed Aircraft Design," *The Journal of the Royal Aeronautical Society*, vol. 39, no. 291, pp. 213–228, 1935, ISSN: 0368-3931. DOI: `10.1017/S0368393100111575`.

[77] Cockburn, A., Masson, D., Gutwin, C., Palanque, P., Goguey, A., Yung, M., Gris, C., and Trask, C., "Design and evaluation of braced touch for touchscreen input stabilisation," *International Journal of Human-Computer Studies*, vol. 122, pp. 21–37, 2019, ISSN: 10715819. DOI: `10.1016/j.ijhcs.2018.08.005`.

[78] Cockburn, A., Gutwin, C., Palanque, P., Deleris, Y., Trask, C., Coveney, A., Yung, M., and MacLean, K., "Turbulent touch: Touchscreen input for cockpit displays," in *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, New York, NY, USA: ACM, 2017, pp. 6742–6753, ISBN: 9781450346559. DOI: `10.1145/3025453.3025584`.

[79] Cohen, J., "A power primer.," *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992, ISSN: 1939-1455. DOI: `10.1037/0033-2909.112.1.155`.

[80] Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*, 2nd. Hillsdale, NJ: Erlbaum, 2013. DOI: `10.4324/9780203771587`.

[81] Cohn, M., *Garmin® Receives EASA Approval for GTN™ Touchscreen Avionics*, 2011. [Online]. Available: `https://www.garmin.com/en-US/newsroom/press-release/aviation/2011-garmin-receives-easa-approval-for-gtn-touchscreen-avionics/` (visited on 12/21/2022).

[82] Colley, A., Henze, N., and Khamis, M., “Touchscreens as the de facto interface to complex systems,” *Managing Complexity and Creating Innovation through Design*, no. April, pp. 89–99, 2019. DOI: `10.4324/9780429022746`.

[83] Comerford, D. and Johnson, W. W., “Potential Capabilities in a Future, Augmented Cockpit,” *Ergonomics in Design: The Quarterly of Human Factors Applications*, vol. 15, no. 1, pp. 8–13, 2007, ISSN: 1064-8046. DOI: `10.1177/106480460701500105`.

[84] Cooley, M., “Human-Centered Design,” in *Information Design*, Jacobson, R., Ed., 1st ed., Cambridge, Mass.: MIT Press, 2000, pp. 59–81, ISBN: 0-262-10069-X.

[85] Cooper, G. E. and Harper, R. P., J., “The use of pilot rating in the evaluation of aircraft handling qualities,” NASA Ames Research Center, Washington, Tech. Rep., 1969.

[86] Correll, J. T., “The Air Mail Fiasco,” *AIR FORCE Magazine*, 2008.

[87] Coutts, L. V., Plant, K. L., Smith, M., Bolton, L., Parnell, K. J., Arnold, J., and Stanton, N. A., “Future technology on the flight deck: assessing the use of touchscreens in vibration environments,” *Ergonomics*, vol. 62, no. 2, pp. 286–304, 2019, ISSN: 13665847. DOI: `10.1080/00140139.2018.1552013`.

[88] Cradle of Aviation Museum and Education Center, *Breese Penguin*, 2022. [Online]. Available: `https://www.cradleofaviation.org/history/exhibits/exhibit-galleries/world%5C_war%5C_i/breese%5C_penguin.html` (visited on 09/15/2022).

[89] Cranfield University, *Flight simulator at Cranfield University wins international award*, 2021. [Online]. Available: `https://www.cranfield.ac.uk/press/news-2021/flight-simulator-at-cranfield-university-wins-international-award` (visited on 10/06/2022).

[90] Cross, J. I., Boag-Hodgson, C., Ryley, T., Mavin, T., and Potter, L. E., “Using Extended Reality in Flight Simulators: A Literature Review,” *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2022, ISSN: 1077-2626. DOI: `10.1109/TVCG.2022.3173921`.

[91] Crossman, E. C., “”Dry Shooting” for Airplane Gunners,” *Popular Science Monthly*, vol. 94, no. 1, pp. 13–14, 1919.

[92] Cushing, S., *Fatal Words: Communication Clashes and Aircraft Crashes.* University of Chicago Press, 1994, ISBN: 0-226-13200-5.

[93] Czaja, S. J. and Sharit, J., "Age differences in the performance of computer-based work.," *Psychology and Aging*, vol. 8, no. 1, pp. 59–67, 1993, ISSN: 1939-1498. DOI: `10.1037/0882-7974.8.1.59`.

[94] Dale L Grover, Martin T King, and Clifford A Kushler, *Reduced Keyboard Disambiguating Computer (patent no US5818437)*, 1998. [Online]. Available: `http://www.google.com/patents?id=PmgCAAAAEBAJ`.

[95] Darrah, D., Moorthamers, B., Anemaat, W., and Liu, W., "Modeling and Optimization of Propulsion Systems for eVTOL Aircraft," in *Proceedings of the Vertical Flight Society 78th Annual Forum*, The Vertical Flight Society, 2022, pp. 1–5. DOI: `10.4050/F-0078-2022-17489`.

[96] Davis, C., *A Closer Look At The Game.Com*, 2001. [Online]. Available: `https://web.archive.org/web/20010709152443/http://www.videogames.com/features/universal/game%5C_com/` (visited on 09/19/2022).

[97] DCA Design International, *Exploring the future of cockpit design*, 2019. [Online]. Available: `https://www.dca-design.com/work/rolls-royce-exploring-future-cockpit-design` (visited on 12/01/2022).

[98] DCA Design International, *Exploring the future of cockpit design*, 2021. [Online]. Available: `https://www.dca-design.com/work/rolls-royce-exploring-future-cockpit-design` (visited on 10/12/2022).

[99] De Maria, G., "Human-machine Interaction in Aviation: A Future Threat or Resource," *American Journal of Science and Technology*, vol. 3, no. 1, pp. 25–42, 2016.

[100] Demaria, R. and Wilson, J. L., *High Score! The Illustrated History of Video games.* McGraw-Hill, 2002, p. 30, ISBN: 978-0-07-222428-3.

[101] Dharanika, T., "Intelligent Wheel Chair for Disabled Person," *International Journal for Innovative Research in Science & Technology*, vol. 3, no. 1, 2016.

[102] Dicke, C., Wolf, K., and Tal, Y., "Foogue: eyes-free interaction for smartphones," in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI '10*, New York, New York, USA: ACM Press, 2010, p. 455, ISBN: 9781605588353. DOI: `10.1145/1851600.1851705`.

[103] Diez, B., “UAM VTOL Flight Control System development & Handling Qualities assessment,” MSc thesis, Cranfield University, 2020.

[104] Digital Nation, *Interviews - P.W. Singer*, 2009. [Online]. Available: `https://www.pbs.org/wgbh/pages/frontline/digitalnation/extras/interviews/singer.html` (visited on 11/07/2022).

[105] Dobrowolski, P., Hanusz, K., Sobczyk, B., Skorko, M., and Wiatrow, A., “Cognitive enhancement in video game players: The role of video game genre,” *Computers in Human Behavior*, vol. 44, pp. 59–63, 2015, ISSN: 07475632. DOI: `10.1016/j.chb.2014.11.051`.

[106] Dodd, S., Lancaster, J., Miranda, A., Grothe, S., DeMers, B., and Rogers, B., “Touch Screens on the Flight Deck: The Impact of Touch Target Size, Spacing, Touch Technology and Turbulence on Pilot Performance,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1, pp. 6–10, 2014, ISSN: 2169-5067. DOI: `10.1177/1541931214581002`.

[107] Dollinger, D., Reiss, P., Angelov, J., Löbl, D., and Holzapfel, F., “Control Inceptor Design for Onboard Piloted Transition VTOL Aircraft Considering Simplified Vehicle Operation,” in *AIAA Scitech 2021 Forum*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2021, pp. 1–11, ISBN: 978-1-62410-609-5. DOI: `10.2514/6.2021-1896`.

[108] Dontre, A. J., “The influence of technology on academic distraction: A review,” *Human Behavior and Emerging Technologies*, vol. 3, no. 3, pp. 379–390, 2021, ISSN: 2578-1863. DOI: `10.1002/hbe2.229`.

[109] Drela, M., “Simultaneous Optimization of the Airframe, Powerplant, and Operation of Transport Aircraft,” in *RAeS 2nd Aircraft Structural Design Conference*, 2010, pp. 1–24.

[110] Dumitru, I. M. and Boşcoianu, M., “Human Factors Contribution To Aviation Safety,” *Scientific Research & Education in the Air Force - AFASES*, vol. 1, pp. 49–53, 2015.

[111] Early, K., “Propulsion airframe integration design, analysis and challenges going into the 21st century,” *The Aeronautical Journal*, vol. 104, no. 1038, pp. 375–382, 2000, ISSN: 0001-9240. DOI: `10.1017/S0001924000064010`.

[112] Eastwood, C., Director, and Komarnicki, T., Producer, *Sully*, [Film]. USA: Warner Bros. Pictures, 2016.

[113] Efremov, A. V., Efremov, E. V., MbiKayi, Z., and Irgaleev, I. K., "Influence of Inceptors on Pilot-Aircraft System Characteristics and Flying Qualities," *IOP Conference Series: Materials Science and Engineering*, vol. 476, no. 1, pp. 1–8, 2019, ISSN: 1757-899X. DOI: `10.1088/1757-899X/476/1/012010`.

[114] Efremov, A. V., Aleksandrov, V. V., Efremov, E. V., and Vukolov, M. V., "The influence of different types of inceptors and their characteristics on the pilot-aircraft system," *IFAC-PapersOnLine*, vol. 51, no. 34, pp. 372–377, 2019, ISSN: 24058963. DOI: `10.1016/j.ifacol.2019.01.013`.

[115] Eichinger, A. and Kellerer, J., "Between laboratory and simulator: a cognitive approach to evaluating cockpit interfaces by manipulating informatory context," *Cognition, Technology & Work*, vol. 16, no. 3, pp. 417–427, 2014, ISSN: 1435-5558. DOI: `10.1007/s10111-013-0270-y`.

[116] Endsley, M. R. and Bolstad, C. A., "Individual Differences in Pilot Situation Awareness," *The International Journal of Aviation Psychology*, vol. 4, no. 3, pp. 241–264, 1994, ISSN: 1050-8414. DOI: `10.1207/s15327108ijap0403_3`.

[117] Engel, J. A., *Cold War at 30,000 Feet: The Anglo-American Fight for Aviation Supremacy*. Harvard University Press, 2007, ISBN: 978-0674024618.

[118] European Aviation Safety Agency, "Certification Specifications for Aeroplane Flight Simulation Training Devices," no. 2, p. 154, 2018.

[119] European Commission, "Technology readiness levels (TRL). Extract from Part 19 - Commission Decision C(2014)4995," Tech. Rep., 2014.

[120] Everitt, B., Landau, S., and Leese, M., *Cluster Analysis*, 4th. London: Arnold, 2001, ISBN: 1-58488-369-3.

[121] Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A., "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior Research Methods*, vol. 39, no. 2, pp. 175–191, 2007, ISSN: 1554-351X. DOI: `10.3758/BF03193146`.

[122] Federal Aviation Administration, *Pilot's Handbook of Aeronautical Knowledge, FAA-H-8083-25B*. 2019, ISBN: 9781602397804.

[123] Fendt, M., *Airbus begins deliveries of first A350 XWBs with touchscreen cockpit displays option to customers*, 2019. [Online]. Available: `https://www.airbus.com/en/newsroom/press-releases/2019-12-airbus-begins-deliveries-of-first-a350-xwbs-with-touchscreen` (visited on 12/21/2022).

[124] Field, A., *Discovering Statistics Using SPSS: (and Sex and Drugs and Rock 'n' Roll)*, Third. London: SAGE Publications Ltd, 2009, ISBN: 978-1-84787-906-6.

[125] Finger, D. F., Braun, C., and Bil, C., "A Review of Configuration Design for Distributed Propulsion Transitioning VTOL Aircraft," *2017 Asia-Pacific International Symposium on Aerospace Technology*, no. October, pp. 1782–1796, 2017.

[126] Finley, P. M., "A study comparing table-based and list-based smartphone interface usability," Master of Fine Arts thesis, Iowa State University, 2013.

[127] Fitts, P. M. and Deininger, R. L., "S-R compatibility: Correspondence among paired elements within stimulus and response codes.," *Journal of Experimental Psychology*, vol. 48, no. 6, pp. 483–492, 1954, ISSN: 0022-1015. DOI: `10.1037/h0054967`.

[128] Fitts, P. M. and Seeger, C. M., "S-R compatibility: spatial characteristics of stimulus and response codes.," *Journal of Experimental Psychology*, vol. 46, no. 3, pp. 199–210, 1953, ISSN: 0022-1015. DOI: `10.1037/h0062827`.

[129] Flaherty, S. R., Fern, L., Turpin, T., and Scheff, S., "Universal Ground Control Station (UGCS) joystick evaluation," in *2012 IEEE Aerospace Conference*, IEEE, 2012, pp. 1–15, ISBN: 978-1-4577-0557-1. DOI: `10.1109/AERO.2012.6187347`.

[130] Flight Safety Foundation, *ASN Wikibase Occurrence # 47464*, 2014. [Online]. Available: `https://aviation-safety.net/wikibase/47464` (visited on 09/16/2022).

[131] Flink, J. J., *The automobile age*. Cambridge, Mass.: MIT Press, 1990, p. 34, ISBN: 978-0262560559.

[132] Garg, A., Linda, R. I., and Chowdhury, T., "Evolution of Aircraft Flight Control System and Fly-By-Light Flight Control System," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 12, pp. 1–6, 2013, ISSN: 2250-2459.

[133] Gary, A., *Axon's Automotive Anorak: That time Saab built a car with a joystick*, 2018. [Online]. Available: `https://www.goodwood.com/grr/road/news/2018/5/axons-automotive-anorak-that-time-saab-built-a-car-with-a-joystick/` (visited on 09/20/2022).

[134] Gauci, J., Cauchi, N., Theuma, K., Zammit-Mangion, D., and Muscat, A., "Design and evaluation of a touch screen concept for pilot interaction with avionic systems," in *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, IEEE, 2015, pp. 3C2–1–3C2–19, ISBN: 978-1-4799-8940-9. DOI: `10.1109/DASC.2015.7311398`.

[135] Ghani, M. A. E., Berthouze, N., Marquardt, N., Jani, R., Hargreaves, D., Wright, S. J., and Smith, H. H., "Tangible Interaction with In-Car Smart Intelligence," in *AutomotiveUI '22: Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, New York, NY, USA: Association for Computing Machinery, 2022, pp. 1–9. DOI: `10.1145/3544999.3552538`.

[136] Gibbs, C. B., "Controller Design: Interactions Of Controlling Limbs, Time-Lags And Gains In Positional And Velocity Systems," *Ergonomics*, vol. 5, no. 2, pp. 385–402, 1962, ISSN: 13665847. DOI: `10.1080/00140136208930602`.

[137] Gil, D., Hernandez-Sabate, A., Enconniere, J., Asmayawati, S., Folch, P., Borrego-Carazo, J., and Piera, M. A., "E-Pilots: A System to Predict Hard Landing During the Approach Phase of Commercial Flights," *IEEE Access*, vol. 10, pp. 7489–7503, 2022, ISSN: 2169-3536. DOI: `10.1109/ACCESS.2021.3138167`.

[138] Gillis, D., Petri, M., Pratelli, A., Semanjski, I., and Semanjski, S., "Urban Air Mobility: A State of Art Analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12950 LNCS, 2021, pp. 411–425, ISBN: 9783030869595. DOI: `10.1007/978-3-030-86960-1_29`.

[139] Gilruth, R. R., "Requirements for satisfactory flying qualities of airplanes," NASA, Tech. Rep., 1943, pp. 49–57.

[140] Girden, E., *ANOVA: Repeated Measures*. Newbury Park, CA: SAGE Publications, Inc., 1992, ISBN: 9780803942578. DOI: `10.4135/9781412983419`.

[141] Golding, R. J., "Flight simulation at Cranfield Institute of Technology," *The Aeronautical Journal*, vol. 84, no. 835, pp. 236–237, 1980. DOI: `10.1017/S0001924000031183`.

[142] Gonçalves, G., Mourão, É., Torok, L., Trevisan, D., Clua, E., and Montenegro, A., "Understanding user experience with game controllers: A case study with an adaptive smart controller and a traditional gamepad," in *Lecture Notes in*

Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10507 LNCS, 2017, pp. 59–71, ISBN: 9783319667140. DOI: `10.1007/978-3-319-66715-7_7`.

[143] Gratton, G., "Crashworthiness and Escape," in *Initial Airworthiness*, Cham: Springer International Publishing, 2018, pp. 213–232. DOI: `10.1007/978-3-319-75617-2_10`.

[144] Grayson, R. L. and Billings, C. E., "Information transfer between air traffic control and aircraft: Communication problems in flight and operations," in *Information Transfer Problems in the Aviation System (NASA TP-1875)*, Billings, C. E. and Cheaney, E., Eds., 1981, pp. 47–62.

[145] Greathouse, J., *Reinvent The Wheel - A Nonstandard Look at Standards*, 2008. [Online]. Available: `https://www.infochachkie.com/wheel/` (visited on 09/20/2022).

[146] Grimaccia, E. and Naccarato, A., "Confirmatory factor analysis to validate a new measure of food insecurity: perceived and actual constructs," *Quality & Quantity*, vol. 54, no. 4, pp. 1211–1232, 2020, ISSN: 0033-5177. DOI: `10.1007/s11135-020-00982-y`.

[147] Gursky, B. I. and Müller, D., "Novel steering concepts for personal aerial vehicle," *Deutscher Luft - und Raumfahrtkongress 2013*, pp. 1–7, 2013.

[148] Haas, E. C. and Kunze, M., "The Effect of a Vehicle Control Device on Driver Performance in a Simulated Tank Driving Task," in *Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design: driving assessment 2001*, Iowa City, Iowa: University of Iowa, 2001, pp. 143–146. DOI: `10.17077/drivingassessment.1025`.

[149] Hancock, P., Vincenzi, D. A., Wise, J. A., and Mouloua, M., *Human Factors in Simulation and Training*, Hancock, P. A., Vincenzi, D. A., Wise, J. A., and Mouloua, M., Eds. CRC Press, 2008, ISBN: 9781420072846. DOI: `10.1201/9781420072846`.

[150] Hanlon, C., Potter, C. C., Wingett, P. T., Wilkens, D. R., and Abel, S. G., *Active pilot flight control stick system with passive electromagnetic feedback*, 2008. [Online]. Available: `https://www.freepatentsonline.com/y2008/0156939.html`.

[151] Harper, R. P. and Cooper, G. E., "Handling qualities and pilot evaluation," *Journal of Guidance, Control, and Dynamics*, vol. 9, no. 5, pp. 515–529, 1986, ISSN: 0731-5090. DOI: `10.2514/3.20142`.

[152] Harris, D., "A human-centred design agenda for the development of single crew operated commercial aircraft," *Aircraft Engineering and Aerospace Technology*, vol. 79, no. 5, pp. 518–526, 2007, ISSN: 00022667. DOI: `10.1108/00022660710780650`.

[153] Harris, D., "The future flight deck," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10276 LNAI, 2017, pp. 222–230, ISBN: 9783319584744. DOI: `10.1007/978-3-319-58475-1_17`.

[154] Hart, S. G., "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006, ISSN: 2169-5067. DOI: `10.1177/154193120605000909`.

[155] Hart, S. G. and Staveland, L. E., "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*, C, vol. 52, 1988, pp. 139–183. DOI: `10.1016/S0166-4115(08)62386-9`.

[156] Hartman, B. O. and Secrist, G. E., "Situational awareness is more than exceptional vision," *Aviation, Space and Environmental Medicine*, vol. 62, no. 11, pp. 1084–1089, 1991.

[157] Havkar.com, *First Aircraft Simulator*, 2018. [Online]. Available: `http://havkar.com/en/blog/view/first-aircraft-simulator/90` (visited on 09/15/2022).

[158] Haward, D. M., "The Sanders Teacher," *Flight*, vol. 2, no. 50, 1910.

[159] Hays, R. T., Jacobs, J. W., Prince, C., and Salas, E., "Flight Simulator Training Effectiveness: A Meta-Analysis," *Military Psychology*, vol. 4, no. 2, pp. 63–74, 1992, ISSN: 0899-5605. DOI: `10.1207/s15327876mp0402_1`.

[160] Héder, M., "From NASA to EU: The evolution of the TRL scale in Public Sector Innovation," *Innovation Journal*, vol. 22, no. 2, pp. 1–23, 2017, ISSN: 17153816.

[161] Hellings, G. and Emms, E., "A visual system for flight simulators," *British Communications and Electronics*, vol. 7, no. 5, pp. 334–337, 1960.

[162] Hill, N. and Guillenchmidt, P., "The possibility of the flight simulator as a training aid to helicopter pilots," *The Journal of the Helicopter Association of Great Britain*, vol. 8, no. 4, pp. 133–157, 1955.

[163] Hinrichsen, J. and Bautista, C., "The challenge of reducing both airframe weight and manufacturing cost," *Air & Space Europe*, vol. 3, no. 3-4, pp. 119–121, 2001, ISSN: 12900958. DOI: `10.1016/S1290-0958(01)90072-3`.

[164] Hinz, M., *Statistical methods for archaeological data analysis I: Basic methods. Cluster Analysis (presentation slides)*, 2019. [Online]. Available: `https://martinhinz.github.io/smada%5C_I%5C_2019/11%5C_session/session%5C_11%5C_cluster%5C_analysis.html%5C#1` (visited on 12/15/2022).

[165] Hodgkinson, J., "History of Low-Order Equivalent Systems for Aircraft Flying Qualities," *Journal of Guidance, Control, and Dynamics*, vol. 28, no. 4, pp. 577–583, 2005, ISSN: 0731-5090. DOI: `10.2514/1.3787`.

[166] Holec, P. V. and Hackel, B. J., "PyMOL360: Multi-user gamepad control of molecular visualization software," *Journal of Computational Chemistry*, vol. 37, no. 30, pp. 2667–2669, 2016, ISSN: 01928651. DOI: `10.1002/jcc.24489`.

[167] Horne, T. A., *A fly-by-wire future? Digital flight controls, on the march*, 2020. [Online]. Available: `https://www.aopa.org/news-and-media/all-news/2020/october/pilot/a-fly-by-wire-future` (visited on 10/05/2022).

[168] Horowitz, K., "Sega's Innovation Conquers U.S. Arcades (1981-1982): Turbo (October 1981)," in *The Sega Arcade Revolution: A History in 62 Games*, McFarland, 2018, pp. 43–46, ISBN: 1476672253.

[169] HP, *HP-150 touchscreen personal computer with HP 9121 dual drives, 1983*, 2012. [Online]. Available: `http://www.hp.com/hpinfo/abouthp/histnfacts/museum/personalsystems/0031/` (visited on 12/03/2022).

[170] Huang, S., *Compute the Kendall's coefficient of concordance of the matrix X*, 2010. [Online]. Available: `https://www.mathworks.com/matlabcentral/fileexchange/27833-compute-the-kendall-s-coefficient-of-concordance-of-the-matrix-x` (visited on 12/16/2022).

[171] IBM, *Bivariate Correlations*, 2021. [Online]. Available: `https://www.ibm.com/docs/en/spss-statistics/27.0.0?topic=features-bivariate-correlations` (visited on 10/22/2022).

[172] In, J., "Introduction of a pilot study," *Korean Journal of Anesthesiology*, vol. 70, no. 6, p. 601, 2017, ISSN: 2005-6419. DOI: `10.4097/kjae.2017.70.6.601`.

[173] International Arcade Museum, *Astro Race*, 2022. (visited on 09/19/2022).

[174] International Telecommunication Union, "ITU-T Recommendation E.161: Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network," International Telecommunication Union, Tech. Rep., 2001.

[175] Irvine, D., Zemke, A., Pusateri, G., Gerlach, L., Chun, R., and Jay, W. M., "Tablet and Smartphone Accessibility Features in the Low Vision Rehabilitation," *Neuro-Ophthalmology*, vol. 38, no. 2, pp. 53–59, 2014, ISSN: 0165-8107. DOI: `10.3109/01658107.2013.874448`.

[176] IvyPanda, *Pilots and Minimum Flight Hours*, 2020. [Online]. Available: `https://ivypanda.com/essays/pilots-and-minimum-flight-hours/` (visited on 11/04/2022).

[177] Jaguar Land Rover, *"Hello, My Name Is Sayer. I Am The Steering Wheel Of The Future"*, 2017. [Online]. Available: `https://media.jaguar.com/news/2017/09/hello-my-name-sayer-i-am-steering-wheel-future` (visited on 09/20/2022).

[178] Jamieson, C., *Remember that time Mercedes literally reinvented the wheel?* 2021. [Online]. Available: `https://www.topgear.com/car-news/concept/remember-time-mercedes-literally-reinvented-wheel` (visited on 09/20/2022).

[179] Jansen, C., "Present and future developments of the NLR moving base research flight simulator," in *Flight Simualtion Technologies Conference*, Reston, Virigina: American Institute of Aeronautics and Astronautics, 1988, pp. 54–61. DOI: `10.2514/6.1988-4584`.

[180] Jensen, A., *Honeywell Technologies Take to the Skies in Gulfstream's Newest Certified Aircraft*, 2018. [Online]. Available: `https://aerospace.honeywell.com/us/en/about-us/press-release/2018/07/honeywell-technologies-take-to-the-skies-in-gulfstreams` (visited on 09/26/2022).

[181] Jensen, L. and Konradsen, F., "A review of the use of virtual reality head-mounted displays in education and training," *Education and Information Technologies*, vol. 23, no. 4, pp. 1515–1529, 2018, ISSN: 1360-2357. DOI: `10.1007/s10639-017-9676-0`.

[182] JJ, *MAE and RMSE — Which Metric is Better?* 2016. [Online]. Available: `https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d` (visited on 05/04/2022).

[183] Johnson, E. A., "Touch display - a novel input/output device for computers," *Electronics Letters*, vol. 1, no. 8, 1965. DOI: `10.1049/el:19650200`.

[184] Johnson, E. A., *Touch Displays*, 1965. [Online]. Available: `https://patents.google.com/patent/GB1172222A/en?inventor=Eric+Arthur+Johnson%5C&sort=old%5C&page=1`.

[185] Jones, F., Passos-Neto, C., and Melro Braghiroli, O., "Simulation in Medical Education: Brief history and methodology," *Principles and Practice of Clinical Research Journal*, vol. 1, no. 2, pp. 56–63, 2015, ISSN: 23781890. DOI: `10.21801/ppcrj.2015.12.8`.

[186] Jones, M., Perfect, P., Jump, M., and White, M., "Investigation of novel concepts for control of a personal air vehicle," *Annual Forum Proceedings - AHS International*, vol. 3, no. January, pp. 2329–2344, 2014, ISSN: 15522938.

[187] Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kim, S.-H., Kaufmann, K., and Hsiang, S., "Perceived Clutter in Advanced Cockpit Displays: Measurement and Modeling with Experienced Pilots," *Aviation, Space, and Environmental Medicine*, vol. 79, no. 11, pp. 1007–1018, 2008, ISSN: 00956562. DOI: `10.3357/ASEM.2319.2008`.

[188] Kelber, C., Webber, D., Gomes, G., Lohmann, M., Rodrigues, M., and Ledur, D., "Active steering unit with integrated ACC for X-by-wire vehicles using a joystick as H.M.I.," in *IEEE Intelligent Vehicles Symposium*, IEEE, 2004, pp. 173–177, ISBN: 0-7803-8310-9. DOI: `10.1109/IVS.2004.1336376`.

[189] Kennedy, Q., Taylor, J. L., Reade, G., and Yesavage, J. A., "Age and Expertise Effects in Aviation Decision Making and Flight Control in a Flight Simulator," *Aviation, Space, and Environmental Medicine*, vol. 81, no. 5, pp. 489–497, 2010, ISSN: 00956562. DOI: `10.3357/ASEM.2684.2010`.

[190] Kiljander, H., "Evolution and Usability of Mobile Phone Interaction Styles," PhD thesis, Helsinki University of Technology, 2004, ISBN: 9512273195.

[191] Kimes, B., *Standard catalog of American Cars 1805–1942*, 3rd ed. Krause publications, 1996, pp. 200–245, ISBN: 0-87341-478-0.

[192] Kivila, A., "Touchscreen interfaces for machine control and education," Master of Science thesis, Georgia Institute of Technology, 2013.

[193] Kluska, B., "Magnavox Odyssey. Poczatek odysei (in Polish)," *CD-Action Retro. Wydanie Specjalne*, vol. 3, pp. 26–27, 2022.

[194] Klyde, D. H., Schulze, P. C., Mitchell, D., and Alexandrov, N., "Development of a Process to Define Unmanned Aircraft Systems Handling Qualities," in *2018 AIAA Atmospheric Flight Mechanics Conference*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2018, ISBN: 978-1-62410-525-8. DOI: `10.2514/6.2018-0299`.

[195] Knaack, M. S., "B-52 Straiofortress," in *Encyclopedia of U.S. Air Force Aircraft and Missile Systems Vol. 2: Post-World War II Bombers*, Washington, D.C.: U.S. Government Printing Office, 1988, ISBN: 0-912799-59-5.

[196] Koon, J., *How Fiber Optics Will Propel Future Avionics*, 2020. [Online]. Available: `https://interactive.aviationtoday.com/how-fiber-optics-will-propel-future-avionics/` (visited on 09/05/2022).

[197] Korek, W. T., Li, W.-C., Lu, L., and Lone, M., "Investigating Pilots' Operational Behaviours While Interacting with Different Types of Inceptors," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13307 LNAI, Springer International Publishing, 2022, pp. 314–325, ISBN: 9783031060854. DOI: `10.1007/978-3-031-06086-1_24`.

[198] Korek, W. T., Mendez, A., Asad, H. U., Li, W.-C., and Lone, M. M., "Understanding Human Behaviour in Flight Operation Using Eye-Tracking Technology," in *Engineering Psychology and Cognitive Ergonomics. Cognition and Design. HCII 2020. Lecture Notes in Computer Science*, Harris, D. and Li, W., Eds., vol. 12187, Springer International Publishing, 2020, pp. 304–320. DOI: `10.1007/978-3-030-49183-3\_24`.

[199] Kornecki, A. J. and Hall, K., "Approaches to assure safety in fly-by-wire systems: Airbus vs. boeing," in *Proceedings of the Eigtht IASTED International Conference on Software Engineering and Applications*, 2004, pp. 471–476, ISBN: 0889864276.

[200] Kortum, P. and Oswald, F. L., "The Impact of Personality on the Subjective Assessment of Usability," *International Journal of Human–Computer Interaction*, vol. 34, no. 2, pp. 177–186, 2018, ISSN: 1044-7318. DOI: `10.1080/10447318.2017.1336317`.

[201] Križanović, I., *Cell phone history: From the first phone to today's smartphone wonders*, 2020. [Online]. Available: `https://versus.com/en/news/cell-phone-history` (visited on 09/22/2022).

[202] Krol, L. R., *Multigradient: custom gradient colormap*, 2021. [Online]. Available: `https://github.com/lrkrol/multigradient` (visited on 12/17/2022).

[203] Kwon, S., Choi, E., and Chung, M. K., "Effect of control-to-display gain and movement direction of information spaces on the usability of navigation on small touch-screen interfaces using tap-n-drag," *International Journal of Industrial Ergonomics*, vol. 41, no. 3, pp. 322–330, 2011, ISSN: 01698141. DOI: `10.1016/j.ergon.2011.02.012`.

[204] Al-Lami, H., Aslam, A., Quigley, T., Lewis, J., Mercer, R., and Shukla, P., "The evolution of flight control systems," University of the West of England, Bristol, Tech. Rep., 2015.

[205] Langewiesche, W., *Stick and Rudder: An Explanation of the Art of Flying.* McGraw-Hill Education, 1990, ISBN: 9780070362406.

[206] Lara, D., Romero, G., Ibarra, R., Sanchez, A., and Pegard, C., "Onboard system for flight control of a small UAV," *World Automation Congress Proceedings*, 2012, ISSN: 21544824.

[207] Large, D. R., Banks, V., Burnett, G., and Margaritis, N., "Putting the Joy in Driving: Investigating the Use of a Joystick as an Alternative to Traditional Controls within Future Autonomous Vehicles," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, New York, NY, USA: ACM, 2017, pp. 31–39, ISBN: 9781450351508. DOI: `10.1145/3122986.3122996`.

[208] Lavelle, T. M., Plencner, R. M., and Seidel, J. A., "Concurrentoptimization of airframe and engine design parameters," in *4th Symposium on Multidisciplinary Analysis and Optimization, 1992*, Reston, Virigina: American Institute of Aeronautics and Astronautics, 1992. DOI: `10.2514/6.1992-4713`.

[209] Lawson, C. P., "Environmental Control Systems," in *Encyclopedia of Aerospace Engineering*, Chichester, UK: John Wiley & Sons, Ltd, 2010. DOI: `10.1002/9780470686652.eae468`.

[210] Leishman, J. G., *A History of Helicopter Flight*, 2008. [Online]. Available: `https://web.archive.org/web/20140713201846/http://terpconnect.umd.edu/~leishman/Aero/history.html` (visited on 09/21/2022).

[211] Letondal, C., Vinot, J. L., Pauchet, S., Boussiron, C., Rey, S., Becquet, V., and Lavenir, C., "Being in the sky: Framing tangible and embodied interaction for future airliner cockpits," in *TEI 2018 - Proceedings of the 12th International Conference on Tangible, Embedded, and Embodied Interaction*, vol. 2018-Janua, New York, NY, USA: ACM, 2018, pp. 656–666, ISBN: 9781450355681. DOI: `10.1145/3173225.3173229`.

[212] Letondal, C., Vinot, J.-L., Pauchet, S., and Leriche, S., "Exploring a touch-based flight control panel for pilots using tangible design principles," *Transportation Research Procedia*, vol. 43, pp. 257–268, 2019, ISSN: 23521465. DOI: `10.1016/j.trpro.2019.12.041`.

[213] Lewis, Z. L., Douglas, G. P., Monaco, V., and Crowley, R. S., "Touchscreen Task Efficiency and Learnability in an Electronic Medical Record at the Point-of-Care," *Studies in Health Technology and Informatics*, vol. 160, pp. 101–105, 2010. DOI: `10.3233/978-1-60750-588-4-101`.

[214] Leyman, C., "A review of the technical development of concorde," *Progress in Aerospace Sciences*, vol. 23, no. 3, pp. 185–238, 1986, ISSN: 03760421. DOI: `10.1016/0376-0421(86)90007-2`.

[215] Li, W.-C., Horn, A., Sun, Z., Zhang, J., and Braithwaite, G., "Augmented visualization cues on primary flight display facilitating pilot's monitoring performance," *International Journal of Human Computer Studies*, vol. 135, p. 102 377, 2020, ISSN: 10959300. DOI: `10.1016/j.ijhcs.2019.102377`.

[216] Li, W.-C., Korek, W. T., Liang, Y. H., and Lin, J. J. H., "Touchscreen Controls for Future Flight Deck Design: Investigating Visual Parameters on Human-Computer Interactions between Pilot Flying and Pilot Monitoring," *Journal of Aeronautics, Astronautics and Aviation*, vol. 55, no. 2, pp. 201–211, 2023. DOI: `10.6125/JoAAA.202306_55(2).08`.

[217] Li, W.-C., Liang, Y. H., Korek, W. T., and Lin, J. J., "Assessments on Human-Computer Interaction Using Touchscreen as Control Inputs in Flight Operations," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13307 LNAI, pp. 326–338, 2022, ISSN: 16113349. DOI: `10.1007/978-3-031-06086-1_25`.

[218] Li, W.-C., Liang, Y.-h., and Korek, W. T., "Flight operations using touchscreen controls: assessing system usability and pilots' visual attention," in *Contemporary Ergonomics & Human Factors 2022: Proceedings for the Annual Conference of the Chartered Institute of Ergonomics and Human Factors*, Balfe, N. and Golightly, D., Eds., Chartered Institute of Ergonomics & Human Factors, 2022, ISBN: 978-1-9996527-4-6.

[219] Li, W.-C., Wang, Y., and Korek, W. T., "To be or not to be? Assessment on using touchscreen as inceptor in flight operation," *Transportation Research Procedia*, vol. 66, pp. 117–124, 2022, ISSN: 23521465. DOI: `10.1016/j.trpro.2022.12.013`.

[220] Li, W.-C., Yan, Z., Zhang, J., Braithwaite, G., Court, S., Lone, M. M., and Thapa, B., "Evaluating pilot's perceived workload on interacting with augmented reality device in flight operations," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12187 LNAI, pp. 332–340, 2020, ISSN: 16113349. DOI: `10.1007/978-3-030-49183-3_26`.

[221] Li, W.-C., Yu, C.-S., Greaves, M., and Braithwaite, G., "How Cockpit Design Impacts Pilots' Attention Distribution and Perceived Workload during Aiming a Stationary Target," *Procedia Manufacturing*, vol. 3, pp. 5663–5669, 2015, ISSN: 23519789. DOI: `10.1016/j.promfg.2015.07.781`.

[222] Li, W.-C., Zhang, J., Braithwaite, G., and Kearney, P., "Quick coherence technique facilitating commercial pilots' psychophysiological resilience to the impact of COVID-19," *Ergonomics*, pp. 1–14, 2022, ISSN: 0014-0139. DOI: `10.1080/00140139.2022.2139416`.

[223] Li, W.-C., Zhang, J., Le Minh, T., Cao, J., and Wang, L., "Visual scan patterns reflect to human-computer interactions on processing different types of messages in the flight deck," *International Journal of Industrial Ergonomics*, vol. 72, pp. 54–60, 2019, ISSN: 01698141. DOI: `10.1016/j.ergon.2019.04.003`.

[224] Lim, Y., Gardi, A., Sabatini, R., Ramasamy, S., Kistan, T., Ezer, N., Vince, J., and Bolia, R., "Avionics Human-Machine Interfaces and Interactions for Manned and Unmanned Aircraft," *Progress in Aerospace Sciences*, vol. 102, pp. 1–46, 2018, ISSN: 03760421. DOI: `10.1016/j.paerosci.2018.05.002`.

[225] Lindner, A. and Greul, R., "Does the steering wheel have a future? A UX study," in *9th International Munich Chassis Symposium 2018*, 2019, pp. 527–535. DOI: `10.1007/978-3-658-22050-1_34`.

[226] Link, E. A., *Combination training device for student aviators and entertainment apparatus. US patent no US1825462A.* 1931.

[227] Liptak, A., *The US Navy's newest submarine comes with an Xbox controller*, 2018. [Online]. Available: `https://www.theverge.com/2018/3/18/17136808/us-navy-uss-colorado-xbox-controller` (visited on 11/07/2022).

[228] Littell, J., "Challenges in Vehicle Safety and Occupant Protection for Autonomous electric Vertical Take-off and Landing (eVTOL) Vehicles," in *AIAA Propulsion and Energy 2019 Forum*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2019, ISBN: 978-1-62410-590-6. DOI: `10.2514/6.2019-4504`.

[229] Logitech, *Extreme 3D Pro Joystick*, 2003. [Online]. Available: `https://www.logitechg.com/en-gb/products/space/extreme-3d-pro-joystick.942-000031.html` (visited on 12/21/2022).

[230] Loguidice, B., *A History of Gaming Platforms: Mattel Intellivision*, 2008. [Online]. Available: `https://www.gamedeveloper.com/console/a-history-of-gaming-platforms-mattel-intellivision` (visited on 09/19/2022).

[231] Lombaerts, T., Kaneshige, J., and Feary, M., "Control Concepts for Simplified Vehicle Operations of a Quadrotor eVTOL Vehicle," in *AIAA AVIATION 2020 FORUM*, vol. 1 PartF, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2020, pp. 1–31, ISBN: 978-1-62410-598-2. DOI: `10.2514/6.2020-3189`.

[232] Lombaerts, T., Schuet, S., Acosta, D. M., Kaneshige, J., and Shish, K. H., "Piloted Simulator Evaluation of Maneuvering Envelope Information for Flight Crew Awareness," in *AIAA Guidance, Navigation, and Control Conference*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2015, pp. 1–24, ISBN: 978-1-62410-339-1. DOI: `10.2514/6.2015-1546`.

[233] Lone, M. M., "Pilot modelling for airframe loads analysis," PhD thesis, Cranfield University, 2013.

[234] Lone, M. M., Ruseno, N., and Cooke, A. K., “Towards understanding effects of non-linear flight control system elements on inexperienced pilots,” *Aeronautical Journal*, vol. 116, no. 1185, pp. 1201–1206, 2012, ISSN: 00019240. DOI: `10.1017/S0001924000007569`.

[235] Luk, P. C.-K., “Superconducting machines — The enabling technology for future electric propulsion in aircraft,” in *2017 7th International Conference on Power Electronics Systems and Applications - Smart Mobility, Power Transfer & Security (PESA)*, IEEE, 2017, pp. 1–7, ISBN: 978-1-5386-1387-0. DOI: `10.1109/PESA.2017.8277766`.

[236] Ma, Z., Zhang, W., Zhou, Y., Yu, J., and Li, B., “The glare evaluation method using digital camera for civil airplane flight deck,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, PART 2, vol. 8020 LNAI, 2013, pp. 184–192, ISBN: 9783642393532. DOI: `10.1007/978-3-642-39354-9_21`.

[237] Mahmoodi, F., Kashani-Vahid, L., Moradi, H., and Yekta-Parast, A., “A Cognitive-Sensory-Motor Gamepad for Therapy of Children with ADHD,” in *2019 International Serious Games Symposium (ISGS)*, IEEE, 2019, pp. 25–29, ISBN: 978-1-7281-6766-4. DOI: `10.1109/ISGS49501.2019.9047013`.

[238] Malfatti, S. M., Dos Santos, F. F., and Dos Santos, S. R., “Using mobile phones to control desktop multiplayer games,” in *Proceedings - 2010 Brazilian Symposium on Games and Digital Entertainment, SBGames 2010*, IEEE, 2010, pp. 230–238, ISBN: 9780769543598. DOI: `10.1109/SBGAMES.2010.32`.

[239] Martinez-Val, R., Roa, J., Perez, E., and Cuerno, C., “Effects of the Mismatch Between Design Capabilities and Actual Aircraft Utilization,” *Journal of Aircraft*, vol. 48, no. 6, pp. 1921–1927, 2011, ISSN: 0021-8669. DOI: `10.2514/1.C031348`.

[240] Martiniello, N., Eisenbarth, W., Lehane, C., Johnson, A., and Wittich, W., “Exploring the use of smartphones and tablets among people with visual impairments: Are mainstream devices replacing the use of traditional visual aids?” *Assistive Technology*, vol. 34, no. 1, pp. 34–45, 2022, ISSN: 1040-0435. DOI: `10.1080/10400435.2019.1682084`.

[241] MathWorks, *Hierarchical Clustering*, 2022. [Online]. Available: `https://uk.mathworks.com/help/stats/hierarchical-clustering.html` (visited on 12/14/2022).

[242] McClumpha, A. J., James, M., Green, R. G., and Belyavin, A. J., “Pilots’ Attitudes to Cockpit Automation,” *Proceedings of the Human Factors Society Annual Meeting*, vol. 35, no. 2, pp. 107–111, 1991, ISSN: 0163-5182. DOI: `10.1518/107118191786755698`.

[243] Mcdermott, J., *Attack submarine Colorado, Xbox controllers included, to join the fleet*, 2018. [Online]. Available: `https://eu.usatoday.com/story/news/nation/2018/03/16/attack-submarine-colorado-xbox-controllers-included-join-fleet/431719002/` (visited on 11/07/2022).

[244] McKavitt Jr, T. P., “Parameter identification studies on the NASA/Ames Research Center Advanced Concepts Flight Simulator,” Engineer’s Thesis, Naval Postgraduate School Monterey, CA, United States, 1990.

[245] Mclellan, S., Muddimer, A., and Peres, S. C., “The Effect of Experience on System Usability Scale Ratings,” *Journal of Usability Studies*, vol. 7, no. 2, pp. 56–67, 2012, ISSN: 1931-3357.

[246] McMillan, D., “Say again? Miscommunications in Air Traffic Control,” Master of Education thesis, Queensland University of Technology, 1998, pp. 1–61.

[247] McRuer, D. and Jex, H., “A Review of Quasi-Linear Pilot Models,” *IEEE Transactions on Human Factors in Electronics*, vol. HFE-8, no. 3, pp. 231–249, 1967, ISSN: 0096-249X. DOI: `10.1109/THFE.1967.234304`.

[248] Mercedes-Benz Group, *1885–1886: The first automobile*, 2022. [Online]. Available: `https://group.mercedes-benz.com/company/tradition/company-history/1885-1886.html` (visited on 09/19/2022).

[249] Merdenyan, B. and Petrie, H., “User reviews of gamepad controllers: A source of user requirements and user experience,” in *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, New York, NY, USA: ACM, 2015, pp. 643–648, ISBN: 9781450334662. DOI: `10.1145/2793107.2810332`.

[250] Merriken, M. S., Johnson, W. V., Cress, J. D., and Riccio, G. E., “Time delay compensation using supplementary cues in aircraft simulator systems,” in *Flight Simualtion Technologies Conference, 1988*, Reston, Virigina: American Institute of Aeronautics and Astronautics, 1988, pp. 295–303. DOI: `10.2514/6.1988-4626`.

[251] Merriman, S. C. and Karn, K. S., "History of Human Factors in US Navy Aircraft Cockpit Design: 1969-2019," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1, pp. 1883–1887, 2019, ISSN: 2169-5067. DOI: `10.1177/1071181319631405`.

[252] Messaoudi, F., Simon, G., and Ksentini, A., "Dissecting games engines: The case of Unity3D," in *2015 International Workshop on Network and Systems Support for Games (NetGames)*, IEEE, 2015, pp. 1–6, ISBN: 978-1-5090-0068-5. DOI: `10.1109/NetGames.2015.7382990`.

[253] Mi, N., Cavuoto, L. A., Benson, K., Smith-Jackson, T., and Nussbaum, M. A., "A heuristic checklist for an accessible smartphone interface design," *Universal Access in the Information Society*, vol. 13, no. 4, pp. 351–365, 2014, ISSN: 1615-5289. DOI: `10.1007/s10209-013-0321-4`.

[254] Mitchell, D. G., "Fifty years of the Cooper-Harper Scale," *AIAA Scitech 2019 Forum*, no. January, pp. 1–18, 2019. DOI: `10.2514/6.2019-0563`.

[255] Mitchell, D. G., Doman, D. B., Key, D. L., Klyde, D. H., Leggett, D. B., Moorhouse, D. J., Mason, D. H., Raney, D. L., and Schmidt, D. K., "Evolution, Revolution, and Challenges of Handling Qualities," *Journal of Guidance, Control, and Dynamics*, vol. 27, no. 1, pp. 12–28, 2004, ISSN: 0731-5090. DOI: `10.2514/1.3252`.

[256] Monahan, S., *Video games have replaced music as the most important aspect of youth culture*, 2021. [Online]. Available: `https://www.theguardian.com/commentisfree/2021/jan/11/video-games-music-youth-culture` (visited on 09/24/2022).

[257] Morgan, R., *Cragstan Periscope-Firing Range*, 2008. [Online]. Available: `https://www.handheldmuseum.com/Misc/CragstanPeriscope.htm` (visited on 09/19/2022).

[258] Moritz, E., Wischgoll, T., and Meyer, J., "Comparison of input devices and displays for protein visualization," *XRDS: Crossroads, The ACM Magazine for Students*, vol. 12, no. 2, pp. 5–5, 2005, ISSN: 1528-4972. DOI: `10.1145/1144375.1144380`.

[259] Mosher, G., *ViewTouch*, 2017. [Online]. Available: `https://www.viewtouch.com/about.html` (visited on 12/03/2022).

[260] Mosquera Benitez, D., del Corte Valiente, A., and Lanzi, P., "A novel global operational concept in cockpits under peak workload situations," *Safety Science*, vol. 102, pp. 38–50, 2018, ISSN: 18791042. DOI: `10.1016/j.ssci.2017.09.028`.

[261] Mundfrom, D. J., Shaw, D. G., and Ke, T. L., "Minimum Sample Size Recommendations for Conducting Factor Analyses," *International Journal of Testing*, vol. 5, no. 2, pp. 159–168, 2005, ISSN: 1530-5058. DOI: `10.1207/s15327574ijt0502_4`.

[262] Murray, H. J. R., *A history of chess*. Northampton, Mass.: Benjamin Press, 1986, ISBN: 0-936317-01-9.

[263] Natapov, D., Castellucci, S. J., and MacKenzie, I. S., "ISO 9241-9 evaluation of video game controllers," in *Proceedings - Graphics Interface*, Kelowna, British Columbia, Canada: Canadian Information Processing Society, 2009, pp. 223–230, ISBN: 9781568814704. DOI: `10.5555/1555880.1555930`.

[264] National Transportation Safety Board, "Aircraft Accident Report - US Airways Flight 1549," *Accident Report*, no. November, p. 213, 2010.

[265] Nieuwenhuizen, F. M., Jump, M., Perfect, P., White, M. D., Padfield, G. D., Floreano, D., Schill, F., Zufferey, J.-C., Fua, P., Bouabdallah, S., Siegwart, R., Meyer, S., Schippl, J., Decker, M., Gursky, B., Höfinger, M., and Bülthoff, H. H., "myCopter: Enabling Technologies for Personal Aerial Transportation Systems," in *3rd International HELI World Conference*, Frankfurt/Main, 2011, pp. 1–8.

[266] Norton, F. H., "The Measurement of the Damping in Roll on a JN4h in Flight," US Government Printing Office, Tech. Rep., 1923.

[267] Nowak, D., Tomczyk, A., and Kopecki, G., "Analysis of the quality control of UAVs for the case of the analytical redundancy of measurements," in *AIAA Modeling and Simulation Technologies Conference*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2016, pp. 1–6, ISBN: 978-1-62410-429-9. DOI: `10.2514/6.2016-3217`.

[268] Nowlan, S., Ebrahimi, A., Whaley, D. R., Demartines, P., Balakrishnan, S., and Rawlins, S., *Data entry apparatus having a limited number of character keys and method (patent no US6204848B1)*, 2001.

[269] O'Regan, G., "Motorola," in *Pillars of Computing*, Cham: Springer International Publishing, 2015, pp. 147–157. DOI: `10.1007/978-3-319-21464-1_23`.

[270] Oberhauser, M. and Dreyer, D., "A virtual reality flight simulator for human factors engineering," *Cognition, Technology and Work*, vol. 19, no. 2-3, pp. 263–277, 2017, ISSN: 14355566. DOI: `10.1007/s10111-017-0421-7`.

[271] Ogren, L. and Wong, J. H., "An analysis of game controller and touchscreen devices for input into a complex high-information display," in *Proceedings of the Human Factors and Ergonomics Society*, vol. 1, 2010, pp. 615–619, ISBN: 9781617820885. DOI: `10.1518/107118110X12829369604163`.

[272] Oliver, T., *The Royal Air Force and the Invention of the Modern Touchscreen*, 2021. [Online]. Available: `https://embeddedcomputing.com/application/consumer/smartphones-and-wearables/the-touch-screen-the-easiest-way-to-keep-in-touch` (visited on 12/03/2022).

[273] Ortega, F. R., Williams, A. S., Tarre, K., Barreto, A., and Rishe, N., "3D Travel Comparison Study between Multi-Touch and GamePad," *International Journal of Human–Computer Interaction*, vol. 36, no. 18, pp. 1699–1713, 2020, ISSN: 1044-7318. DOI: `10.1080/10447318.2020.1780016`.

[274] Oshita, M. and Ishikawa, H., "Gamepad vs. touchscreen: A comparison of action selection interfaces in computer games," in *Proceedings - WASA 2012: Workshop at SIGGRAPH Asia 2012*, New York, New York, USA: ACM Press, 2012, pp. 27–31, ISBN: 9781450318358. DOI: `10.1145/2425296.2425301`.

[275] Page, R., "Brief history of flight simulation," in *SimTecT 2000 Proceedings*, 2000, pp. 1–11.

[276] Paine, L., *The Sea and Civilization: A Maritime History of the World.* London: Atlantic Books, 2014, ISBN: 978-1-78239-357-3.

[277] Palanque, P., Cockburn, A., Désert-Legendre, L., Gutwin, C., and Deleris, Y., "Brace touch: A dependable, turbulence-tolerant, multi-touch interaction technique for interactive cockpits," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11698 LNCS, 2019, pp. 53–68, ISBN: 9783030266004. DOI: `10.1007/978-3-030-26601-1_4`.

[278] Parhi, P., Karlson, A. K., and Bederson, B. B., "Target size study for one-handed thumb use on small touchscreen devices," in *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services - MobileHCI*

'06, New York, New York, USA: ACM Press, 2006, p. 203, ISBN: 1595933905. DOI: `10.1145/1152215.1152260`.

[279] Parnell, K. J., Wynne, R. A., Plant, K. L., Banks, V. A., Griffin, Thomas, G., and Stanton, N. A., "Pilot decision-making during a dual engine failure on take-off: Insights from three different decision-making models," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 32, no. 3, pp. 268–285, 2022, ISSN: 1090-8471. DOI: `10.1002/hfm.20944`.

[280] Pauchet, S., Vinot, J. L., Letondal, C., Lemort, A., Lavenir, C., Lecomte, T., Rey, S., Becquet, V., and Crouzet, G., "Multi-plié: A linear foldable and flattenable interactive display to support efficiency, safety and collaboration exploring and ironing out design complexities with airliner pilots," in *Conference on Human Factors in Computing Systems - Proceedings*, New York, NY, USA: ACM, 2019, pp. 1–13, ISBN: 9781450359702. DOI: `10.1145/3290605.3300384`.

[281] Peacock, C. A., Weber, R., Sanders, G. J., Seo, Y., Kean, D., Pollock, B. S., Burns, K. J., Cain, M., LaScola, P., and Glickman, E. L., "Pilot physiology, cognition and flight performance during flight simulation exposed to a 3810-m hypoxic condition," *International Journal of Occupational Safety and Ergonomics*, vol. 23, no. 1, pp. 44–49, 2017, ISSN: 1080-3548. DOI: `10.1080/10803548.2016.1234685`.

[282] Perry, D. H. and Naish, J. M., "Flight Simulation for Research," *The Journal of the Royal Aeronautical Society*, vol. 68, no. 646, pp. 645–652, 1964, ISSN: 0368-3931. DOI: `10.1017/S0368393100080597`.

[283] Pontius, R. G., Thontteh, O., and Chen, H., "Components of information for multiple resolution comparison between maps that share a real variable," *Environmental and Ecological Statistics*, vol. 15, no. 2, pp. 111–142, 2008, ISSN: 1352-8505. DOI: `10.1007/s10651-007-0043-y`.

[284] Pornet, C. and Isikveren, A., "Conceptual design of hybrid-electric transport aircraft," *Progress in Aerospace Sciences*, vol. 79, pp. 114–135, 2015, ISSN: 03760421. DOI: `10.1016/j.paerosci.2015.09.002`.

[285] Potter, R. L., Weldon, L. J., and Shneiderman, B., "Improving the accuracy of touch screens: an experimental evaluation of three strategies," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '88*, New York, New York, USA: ACM Press, 1988, pp. 27–32, ISBN: 0201142376. DOI: `10.1145/57167.57171`.

[286] Pupil Labs, *Pupil Labs — Core.* 2019. [Online]. Available: `https://pupil-labs.com/products/core/` (visited on 12/09/2019).

[287] Rabin, S., "Nicolaus Copernicus," in *The Stanford Encyclopedia of Philosophy*, Zalta, E. N., Ed., Fall 2019, Metaphysics Research Lab, Stanford University, 2019.

[288] Rajput, N. Z., Khan, A., Kamran, S., Ilyas, M. F., and Ali, S. A., "A Comparative Study of Interface Design of Smartphones," Tech. Rep., 2015, pp. 1–8.

[289] Ram, S., Mahadevan, A., Rahmat-Khah, H., Turini, G., and Young, J. G., "Effect of Control-Display Gain and Mapping and Use of Armrests on Accuracy in Temporally Limited Touchless Gestural Steering Tasks," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 380–384, 2017, ISSN: 2169-5067. DOI: `10.1177/1541931213601577`.

[290] Reid, A. J., "A Brief History of the Smartphone," in *The Smartphone Paradox*, Cham: Springer International Publishing, 2018, pp. 35–66. DOI: `10.1007/978-3-319-94319-0_2`.

[291] Ren, X. and Moriya, S., "Improving selection performance on pen-based systems," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 3, pp. 384–416, 2000, ISSN: 1073-0516. DOI: `10.1145/355324.355328`.

[292] Riccio, G. E., Cress, J. D., and Johnson, W. V., "The Effects of Simulator Delays on the Acquisition of Flight Control Skills: Control of Heading and Altitude," *Proceedings of the Human Factors Society Annual Meeting*, vol. 31, no. 11, pp. 1286–1290, 1987, ISSN: 0163-5182. DOI: `10.1177/154193128703101124`.

[293] Ringham, G. B. and Cutler, A. E., "Flight Simulators," *The Journal of the Royal Aeronautical Society*, vol. 58, no. 519, pp. 153–172, 1954, ISSN: 0368-3931. DOI: `10.1017/S0001924000097918`.

[294] Rolfe, J. and Staples, K., *Flight simulation.* Cambridge University Press, 1986, ISBN: 0521306493.

[295] Roscoe, A. H. and Diffey, B. L., "A Preliminary Study of Blue Light on an Aircraft Flight Deck," *Health Physics*, vol. 66, no. 5, pp. 565–567, 1994, ISSN: 0017-9078. DOI: `10.1097/00004032-199405000-00009`.

[296] Rosen, K. R., “The history of medical simulation,” *Journal of Critical Care*, vol. 23, no. 2, pp. 157–166, 2008, ISSN: 08839441. DOI: `10.1016/j.jcrc.2007.12.004`.

[297] Rothfeld, R., Fu, M., Balać, M., and Antoniou, C., “Potential Urban Air Mobility Travel Time Savings: An Exploratory Analysis of Munich, Paris, and San Francisco,” *Sustainability*, vol. 13, no. 4, p. 2217, 2021, ISSN: 2071-1050. DOI: `10.3390/su13042217`.

[298] Rothfeld, R., Straubinger, A., Fu, M., Al Haddad, C., and Antoniou, C., “Urban air mobility,” in *Demand for Emerging Transportation Systems*, Elsevier, 2020, pp. 267–284. DOI: `10.1016/B978-0-12-815018-4.00013-9`.

[299] Rousseau, R., Tremblay, S., Banbury, S., Breton, R., and Guitouni, A., “The role of metacognition in the relationship between objective and subjective measures of situation awareness,” *Theoretical Issues in Ergonomics Science*, vol. 11, no. 1-2, pp. 119–130, 2010, ISSN: 1463-922X. DOI: `10.1080/14639220903010076`.

[300] Rouwhorst, W., Verhoeven, R., Suijkerbuijk, M., Bos, T., Maij, A., Vermaat, M., and Arents, R., “Use of touch screen display applications for aircraft flight control,” in *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, IEEE, 2017, pp. 1–10, ISBN: 978-1-5386-0365-9. DOI: `10.1109/DASC.2017.8102060`.

[301] Roy, G., *Ship Steering Wheel History*, 2010. [Online]. Available: `https://www.articlesfactory.com/articles/hobbies/ship-steering-wheel-history.html` (visited on 09/20/2022).

[302] Royal Aeronautical Society, *Careers in Aerospace LIVE 2017 - visitor booklet*, 2017. [Online]. Available: `https://www.careersinaerospace.com/wp-content/uploads/2017/11/RAeS%5C_CIAL17%5C_VB%5C_DIGITAL%5C_NOMARKS.pdf` (visited on 12/21/2022).

[303] Rupp, M. A., Oppold, P., and McConnell, D. S., “Comparing the Performance, Workload, and Usability of a Gamepad and Joystick in a Complex Task,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, pp. 1775–1779, 2013, ISSN: 2169-5067. DOI: `10.1177/1541931213571398`.

[304] Sadasivan, S., Vembar, D., Stringfellow, P., Duchowski, A., and Gramopadhye, A., “Evaluation of Interaction Devices for Ndi Training in Vr: Gamepad Vs. Joystick,” *Proceedings of the Human Factors and Ergonomics Society Annual*

Meeting, vol. 50, no. 17, pp. 1995–1999, 2006, ISSN: 2169-5067. DOI: `10.1177/154193120605001760`.

[305] Sadraey, M. H., "Design of control surfaces.," in *Aircraft Design: A Systems Engineering Approach*, Wiley, 2012, ISBN: 978-1-119-95340-1.

[306] Safi, M., Chung, J., and Pradhan, P., "Review of augmented reality in aerospace industry," *Aircraft Engineering and Aerospace Technology*, vol. 91, no. 9, pp. 1187–1194, 2019, ISSN: 1748-8842. DOI: `10.1108/AEAT-09-2018-0241`.

[307] Sager, I., *Before IPhone and Android Came Simon, the First Smartphone*, 2012. [Online]. Available: `https://www.bloomberg.com/news/articles/2012-06-29/before-iphone-and-android-came-simon-the-first-smartphone` (visited on 09/22/2022).

[308] Salman, H. M., Wan Ahmad, W. F., and Sulaiman, S., "Usability Evaluation of the Smartphone User Interface in Supporting Elderly Users From Experts' Perspective," *IEEE Access*, vol. 6, pp. 22 578–22 591, 2018, ISSN: 2169-3536. DOI: `10.1109/ACCESS.2018.2827358`.

[309] Salvendy, G., *Handbook of Human Factors and Ergonomics*, 4th ed., Salvendy, G., Ed. Wiley, 2012, ISBN: 9780470528389. DOI: `10.1002/9781118131350`.

[310] Samsung, *Samsung Galaxy Flip vs Fold: What's The Best Foldable Phone For You*, 2023. [Online]. Available: `https://www.samsung.com/uk/mobile-phone-buying-guide/samsung-fold-vs-flip/` (visited on 01/05/2023).

[311] Sanders, M. and McCormick, E., "Human Factors in Engineering and Design," *Industrial Robot: An International Journal*, vol. 25, no. 2, pp. 153–153, 1998, ISSN: 0143-991X. DOI: `10.1108/ir.1998.25.2.153.2`.

[312] Sarter, N. B. and Woods, D. D., "How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 5–19, 1995, ISSN: 0018-7208. DOI: `10.1518/001872095779049516`.

[313] Saupp, L., Schwarz, B., Schwalm, M., and Eckstein, L., "Evaluation of active Drivesticks as alternative controls in a high fidelity driving simulator," *Automotive and Engine Technology*, vol. 4, no. 1-2, pp. 37–44, 2019, ISSN: 2365-5127. DOI: `10.1007/s41104-019-00039-9`.

[314] Saupp, L., Schwarz, B., Schwalm, M., and Eckstein, L., “Evaluation Of Active Drivesticks As Alternative Controls In Real Traffic,” in *Automotive and Engine Technology*, vol. 4, Aachen, 2019, pp. 37–44. DOI: `10.1007/s41104-019-00039-9`.

[315] Sauro, J., *5 Ways to Interpret a SUS Score*, 2018. [Online]. Available: `https://measuringu.com/interpret-sus-score/` (visited on 12/20/2022).

[316] Schaefer, K. E. and Straub, E. R., “Will passengers trust driverless vehicles? Removing the steering wheel and pedals,” in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, IEEE, 2016, pp. 159–165, ISBN: 978-1-5090-0632-8. DOI: `10.1109/COGSIMA.2016.7497804`.

[317] Schuchardt, B. I., “Workload reduction through steering wheel control for rotorcraft,” *CEAS Aeronautical Journal*, vol. 10, no. 3, pp. 893–902, 2019, ISSN: 18695590. DOI: `10.1007/s13272-018-00360-3`.

[318] Scott-Hall, S., “Aeroplane Performance Testing,” *Aircraft Engineering and Aerospace Technology*, vol. 4, no. 5, pp. 112–114, 1932, ISSN: 0002-2667. DOI: `10.1108/eb029542`.

[319] See, A., “Thrustmaster Formula T1 Driving Simulator Controls by Thrustmaster,” *Game Bytes Magazine*, no. 21, 1994.

[320] Senjam, S., “The current advances in human–smartphone user interface design: An opportunity for people with vision loss,” *Indian Journal of Ophthalmology*, vol. 69, no. 9, p. 2544, 2021, ISSN: 0301-4738. DOI: `10.4103/ijo.IJO_835_21`.

[321] Şenol, M. B., “Anthropometric evaluation of cockpit designs,” *International Journal of Occupational Safety and Ergonomics*, vol. 22, no. 2, pp. 246–256, 2016, ISSN: 1080-3548. DOI: `10.1080/10803548.2015.1126456`.

[322] Sethi, A. K., “Eric Magnus Campbell Tigerstedt,” in *The European Edisons*, New York: Palgrave Macmillan US, 2016, pp. 83–112. DOI: `10.1057/978-1-137-49222-7_3`.

[323] Sfetcu, N., *Game Preview*. Nicolae Sfetcu, 2014, p. 825, ISBN: n/a.

[324] Sherry, L., Polson, P., and Feary, M., “Designing user-interfaces for the cockpit: Five common design errors and how to avoid them,” in *SAE Technical Papers*, 2002. DOI: `10.4271/2002-01-2968`.

[325] Shin, G. and Zhu, X., “User discomfort, work posture and muscle activity while using a touchscreen in a desktop PC setting,” *Ergonomics*, vol. 54, no. 8, pp. 733–744, 2011, ISSN: 0014-0139. DOI: `10.1080/00140139.2011.592604`.

[326] Silfverberg, M., MacKenzie, I. S., and Korhonen, P., “Predicting text entry speed on mobile phones,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’00*, New York, New York, USA: ACM Press, 2000, pp. 9–16, ISBN: 1581132166. DOI: `10.1145/332040.332044`.

[327] Skans, N. S. and Barnes, A. G., “Fifty years of success and failure in flight simulation,” in *50 Years of Flight Simulation, Conference Proceedings, Session 1*, London: The Royal Aeronautical Society, 1979, pp. 33–49.

[328] Snowden, C., “Casting a Powerful Spell: The Evolution of SMS,” in *The Cell Phone Reader: Essays in Social Transformation*, Kavoori, A. P. and Arceneaux, N., Eds., New York: Peter Lang, 2006, pp. 107–108, ISBN: 978-0-8204-7919-4.

[329] Socha, V., Socha, L., Hanakova, L., Lalis, A., Koblen, I., Kusmirek, S., Mrazek, P., Sousek, R., and Schlenker, J., “Basic Piloting Technique Error Rate as an Indicator of Flight Simulators Usability for Pilot Training,” *International Review of Aerospace Engineering (IREASE)*, vol. 9, no. 5, p. 162, 2016, ISSN: 1973-7440. DOI: `10.15866/irease.v9i5.10749`.

[330] Sparmann, E., “’Der Sparmannsche Schulapparat,” *HP-Fachzeitung für Automobilismus und Flugtechnik*, vol. 5, no. 46, pp. 27–29, 1911.

[331] Stanton, N. A., Harvey, C., Plant, K. L., and Bolton, L., “To twist, roll, stroke or poke? A study of input devices for menu navigation in the cockpit,” *Ergonomics*, vol. 56, no. 4, pp. 590–611, 2013, ISSN: 0014-0139. DOI: `10.1080/00140139.2012.751458`.

[332] Statista, *Feature Phones - Worldwide*, 2022. [Online]. Available: `https://www.statista.com/outlook/cmo/consumer-electronics/telephony/feature-phones/worldwide` (visited on 09/23/2022).

[333] Statista, *Smartphones - Worldwide*, 2022. [Online]. Available: `https://www.statista.com/outlook/cmo/consumer-electronics/telephony/smartphones/worldwide` (visited on 09/23/2022).

[334] Steinbock, D. and Wilson, J. L., *The Mobile Revolution.* Kogan Page, 2007, p. 150, ISBN: 978-0-7494-4850-9.

[335] Sullenberger III, C. B. and Zaslow, J., *Highest Duty: My Search for What Really Matters*. William Morrow Paperbacks, 2009, ISBN: 978-0-06-192468-2.

[336] Summers, L. G., Shannon, J. H., White, T. R., and Shiner, R. J., "Fly-by-Wire Sidestick Controller Evaluation," in *SAE Technical Papers*, 1987, pp. 1–14. DOI: `10.4271/871761`.

[337] Takahashi, T., "Ipad's in the Cockpit: Evolution or Revolution in the Sky," *SSRN Electronic Journal*, 2012, ISSN: 1556-5068. DOI: `10.2139/ssrn.2035743`.

[338] Takeda, K., Newman, S. J., Kenny, J., and Zyskowski, M., "Convergence: Commodity flight simulation and the future," *Aeronautical Journal*, vol. 112, no. 1136, pp. 599–607, 2008, ISSN: 00019240. DOI: `10.1017/S0001924000002566`.

[339] Tamsin, M. and Bach, C., "The Design of Medical Devices," *International Journal of Innovation and Scientific Research*, vol. 1, no. 2, pp. 127–134, 2014. DOI: `10.1.1.683.1898`.

[340] Tang, Z. and Zhang, A., "Human-Machine-Cooperation Design Methodology for Civil Aircraft Cockpit," in *Proceedings of the 2nd International Symposium on Computer, Communication, Control and Automation. Advances in Intelligent Systems Research*, 2013. DOI: `10.2991/3ca-13.2013.28`.

[341] Tao, D., Yuan, J., Liu, S., and Qu, X., "Effects of button design characteristics on performance and perceptions of touchscreen use," *International Journal of Industrial Ergonomics*, vol. 64, pp. 59–68, 2018, ISSN: 01698141. DOI: `10.1016/j.ergon.2017.12.001`.

[342] Tao, D., Zeng, J., Liu, K., and Qu, X., "Effects of control-to-display gain and operation precision requirement on touchscreen operations in vibration environments," *Applied Ergonomics*, vol. 91, p. 103 293, 2021, ISSN: 00036870. DOI: `10.1016/j.apergo.2020.103293`.

[343] Taylor, A. K. and Cotter, T. S., "Do Age and Experience Level Affect Views of Pilots' Towards Cockpit Automation," in *Advances in Intelligent Systems and Computing*, vol. 592, 2018, pp. 303–313, ISBN: 9783319603650. DOI: `10.1007/978-3-319-60366-7_29`.

[344] Taylor, R., "Situational Awareness Rating Technique (Sart): The Development of a Tool for Aircrew Systems Design," in *Situational Awareness*, Routledge, 2017, pp. 111–128. DOI: `10.4324/9781315087924-8`.

[345] Telfer, A., *Touch Control Design: Ways of Playing on Mobile*, 2021. [Online]. Available: `https://mobilefreetoplay.com/control-mechanics/` (visited on 06/30/2021).

[346] Thackray, R. I. and Touchstone, R. M., “Age-Related Differences in complex Monitoring Performance,” Federal Aviation Administration, Oklahoma City, Tech. Rep., 1981. DOI: `10.1037/e518802009-001`.

[347] Thales Avionics, *What is new on Avionics 2020?* 2015. [Online]. Available: `https://onboard.thalesgroup.com/new-avionics-2020/` (visited on 09/26/2022).

[348] Thales Avionics, *Airbus Helicopters and French defence procurement agency select Thales's new FlytX avionics suite for latest-generation helicopter programmes*, 2019. [Online]. Available: `https://onboard.thalesgroup.com/airbus-helicopters-and-french-defence-procurement-agency-select-thaless-new-flytx-avionics-suite-for-latest-generation-helicopter-programmes/` (visited on 09/26/2022).

[349] The Drents Museum, *The Pesse canoe*, 2016. [Online]. Available: `https://drentsmuseum.nl/en/in-the-spotlight-top-exhibits/pesse-canoe` (visited on 09/20/2022).

[350] Thomas, P., Biswas, P., and Langdon, P., “State-of-the-Art and Future Concepts for Interaction in Aircraft Cockpits,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9176, 2015, pp. 538–549, ISBN: 9783319206806. DOI: `10.1007/978-3-319-20681-3_51`.

[351] Thrustmaster, *HOTAS WARTHOG™*, 2010. [Online]. Available: `https://www.thrustmaster.com/en-gb/products/hotas-warthog/?platformId=1455` (visited on 12/21/2022).

[352] Thurber, M., *Pilot Report: Gulfstream G500*, 2018. [Online]. Available: `https://www.ainonline.com/aviation-news/business-aviation/2018-08-24/pilot-report-gulfstream-g500` (visited on 12/23/2022).

[353] Tian, F., Chai, W., Wang, C., and Sun, X., “Design and Implementation of Flight Visual Simulation System,” *International Journal of Computer Science Issues*, vol. 9, no. 5, 2012. arXiv: `1212.0365`.

[354] Todd, M. A. and Thomas, M. J., "ATSB Transport Safety Report. Aviation Research Investigation (AR-2012-023)," Australian Transport Safety Bureau, Tech. Rep., 2013.

[355] Tomczyk, A., "Experimental Fly-By-Wire Control System for General Aviation Aircraft," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Reston, Virigina: American Institute of Aeronautics and Astronautics, 2003, pp. 1–8, ISBN: 978-1-62410-090-1. DOI: `10.2514/6.2003-5776`.

[356] Tomczyk, A., "Synthesis of the unmanned aerial vehicle remote control augmentation system," in *AIP Conference Proceedings*, vol. 1637, 2014, pp. 1092–1099, ISBN: 9780735412767. DOI: `10.1063/1.4904684`.

[357] Tomlinson, B. N., "Motion software for a research flight simulator," in *Advances in Flight Simulation - Visual and Motion Systems, International Conference Proceedings*, London: The Royal Aeronautical Society, 1986, pp. 64–80.

[358] Torok, L., Pelegrino, M., Trevisan, D. G., Clua, E., and Montenegro, A., "A Mobile Game Controller Adapted to the Gameplay and User's Behavior Using Machine Learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9353, 2015, pp. 3–16. DOI: `10.1007/978-3-319-24589-8_1`.

[359] Torrance, J., *Global smartphone sales overtake feature phones*, 2013. [Online]. Available: `https://realbusiness.co.uk/global-smartphone-sales-overtake-feature-phones` (visited on 09/22/2022).

[360] ToUsIf, *What is the difference between active and passive sidestick?* 2014. [Online]. Available: `https://aviation.stackexchange.com/questions/5137/what-is-the-difference-between-active-and-passive-sidestick` (visited on 06/16/2020).

[361] Traeger, P., *Gaming as a Cultural Force Just Stepped to the Forefront of Entertainment*, 2020. [Online]. Available: `https://www.adweek.com/sponsored/gaming-as-a-cultural-force-just-stepped-to-the-forefront-of-entertainment/` (visited on 09/24/2022).

[362] Tran, T. H., Behrend, F., Funning, N., and Arango, A., "Single Pilot Operations with AR-Glasses using Microsoft HoloLens," in *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, IEEE, 2018, pp. 1–7, ISBN: 978-1-5386-4112-5. DOI: `10.1109/DASC.2018.8569261`.

[363] Tutubalin, P. I. and Mokshin, V. V., "Structural and Functional Model of the On-board Expert Control System for a Prospective Unmanned Aerial Vehicle," in *Lecture Notes in Electrical Engineering*, vol. 641 LNEE, 2020, pp. 262–272, ISBN: 9783030392246. DOI: `10.1007/978-3-030-39225-3_29`.

[364] Udo-Imeh, N. E. and Landry, S. J., "Dimensions of Pilot Experience and Their Contributing Variables," in *63rd International Symposium on Aviation Psychology*, 2021, pp. 376–384.

[365] Vaart, J. van der, "Modelling of perception and action in compensatory manual control tasks," PhD thesis, Delft University of Technology, 1992.

[366] Vaucher, J., *History of Ships: Prehistoric Craft*, 2014. [Online]. Available: `http://www.iro.umontreal.ca/~vaucher/History/Ships/Prehistoric%5C_Craft/%5C#index` (visited on 09/20/2022).

[367] Verma, W. C. N., Shihab, R. H., and Tobassum, A., "Review Article On History Of Evolution In Cockpit Instruments & Emergence Of All Glass Cockpit," *MIST Journal of Science and Technology*, vol. 2, no. 2, 2013. DOI: `https://doi.org/10.47981/j.mijst.02(02)2013.93(\%25p)`.

[368] Vertegaal, R. and Poupyrev, I., "Organic user Interfaces," *Communications of the ACM*, vol. 51, no. 6, pp. 26–30, 2008, ISSN: 00010782. DOI: `10.1145/1349026.1349033`.

[369] Viertler, F. and Hajek, M., "Requirements and Design Challenges in Rotorcraft Flight Simulations for Research Applications," in *AIAA Modeling and Simulation Technologies Conference*, Reston, Virginia: American Institute of Aeronautics and Astronautics, 2015, ISBN: 978-1-62410-337-7. DOI: `10.2514/6.2015-1808`.

[370] Vincent, B. L., *The Evolution of the Video Game Controller*, 2020. [Online]. Available: `https://www.popularmechanics.com/technology/gadgets/g34288261/evolution-of-the-video-game-controller/` (visited on 09/19/2022).

[371] Virtanen, K., Mansikka, H., Kontio, H., and Harris, D., "Weight watchers: NASA-TLX weights revisited," *Theoretical Issues in Ergonomics Science*, vol. 23, no. 6, pp. 725–748, 2022, ISSN: 1463-922X. DOI: `10.1080/1463922X.2021.2000667`.

[372] Visionary Training Resources, *How it works*, 2022. [Online]. Available: `https://www.vtrvr.com/how-it-works` (visited on 09/16/2022).

[373] Voogt, A. de, “The Transmission of Helicopter Technology, 1920–1939: Exchanges with von Baumhauer,” *The International Journal for the History of Engineering & Technology*, vol. 83, no. 1, pp. 119–140, 2013, ISSN: 1758-1206. DOI: `10.1179/1758120612Z.00000000022`.

[374] Wagner, M., Avdic, D., and Heß, P., “Gamepad Control for Industrial Robots - New Ideas for the Improvement of Existing Control Devices,” in *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics*, SCITEPRESS - Science, 2016, pp. 368–373, ISBN: 978-989-758-198-4. DOI: `10.5220/0005982703680373`.

[375] Wang, L. and Zhang, J., “The effect of psychological risk elements on pilot flight operational performance,” *Human Factors and Ergonomics In Manufacturing*, vol. 30, no. 1, pp. 3–13, 2020, ISSN: 15206564. DOI: `10.1002/hfm.20816`.

[376] Wang, X., Lu, Y., Wang, D., Liu, L., and Zhou, H., “Using Jaccard Distance Measure for Unsupervised Activity Recognition with Smartphone Accelerometers,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10612 LNCS, 2017, pp. 74–83, ISBN: 9783319697802. DOI: `10.1007/978-3-319-69781-9_8`.

[377] Wang, X. and Trasbot, J., “Effects of target location, stature and hand grip type on in-vehicle reach discomfort,” *Ergonomics*, vol. 54, no. 5, pp. 466–476, 2011, ISSN: 0014-0139. DOI: `10.1080/00140139.2011.564312`.

[378] Watkins, C. B., Nilson, C., Taylor, S., Medin, K. B., Kuljanin, I., and Nguyen, H. B., “Development of Touchscreen Displays for the Gulfstream G500 and G600 Symmetry™ Flight Deck,” in *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, IEEE, 2018, pp. 1–10, ISBN: 978-1-5386-4112-5. DOI: `10.1109/DASC.2018.8569532`.

[379] Weick, K. E., “The Vulnerable System: An Analysis of the Tenerife Air Disaster,” *Journal of Management*, vol. 16, no. 3, pp. 571–593, 1990, ISSN: 0149-2063. DOI: `10.1177/014920639001600304`.

[380] Wheeler, P., Sirimanna, T. S., Bozhko, S., and Haran, K. S., “Electric/Hybrid-Electric Aircraft Propulsion Systems,” *Proceedings of the IEEE*, vol. 109, no. 6, pp. 1115–1127, 2021, ISSN: 0018-9219. DOI: `10.1109/JPROC.2021.3073291`.

[381] Wickens, C. D., “Situation Awareness and Workload in Aviation,” *Current Directions in Psychological Science*, vol. 11, no. 4, pp. 128–133, 2002, ISSN: 0963-7214. DOI: `10.1111/1467-8721.00184`.

[382] Wikipedia contributors, *Flight simulator*, 2022. [Online]. Available: `https://en.wikipedia.org/wiki/Flight%5C_simulator` (visited on 09/15/2022).

[383] Wikipedia contributors, *Handheld game console*, 2022. [Online]. Available: `https://en.wikipedia.org/wiki/Handheld%5C_game%5C_console` (visited on 09/19/2022).

[384] Wikipedia contributors, *Joystick*, 2022. [Online]. Available: `https://en.wikipedia.org/wiki/Joystick` (visited on 09/19/2022).

[385] Wikipedia contributors, *Mobile phone*, 2022. [Online]. Available: `https://en.wikipedia.org/wiki/Mobile%5C_phone` (visited on 09/22/2022).

[386] Willmott, C. J., Robeson, S. M., and Matsuura, K., “A refined index of model performance,” *International Journal of Climatology*, vol. 32, no. 13, pp. 2088–2094, 2012, ISSN: 08998418. DOI: `10.1002/joc.2419`.

[387] Winter, J. C. F. de, Dodou, D., and Mulder, M., “Training Effectiveness of Whole Body Flight Simulator Motion: A Comprehensive Meta-Analysis,” *The International Journal of Aviation Psychology*, vol. 22, no. 2, pp. 164–183, 2012, ISSN: 1050-8414. DOI: `10.1080/10508414.2012.663247`.

[388] Witkowski, W., *Videogames are a bigger industry than movies and North American sports combined, thanks to the pandemic*, 2020. [Online]. Available: `https://www.marketwatch.com/story/videogames-are-a-bigger-industry-than-sports-and-movies-combined-thanks-to-the-pandemic-11608654990` (visited on 09/24/2022).

[389] Wolfert, F., Bromfield, M. A., Stedmon, A., and Scott, S., “Passive sidesticks and hard landings – is there a link?” *AIAA Aviation 2019 Forum*, pp. 1–10, 2019. DOI: `10.2514/6.2019-3611`.

[390] Woods, D. D., Dekker, S., Cook, R., Johannesen, L., and Sarter, N., *Behind Human Error*. CRC Press, 2017, ISBN: 9781315568935. DOI: `10.1201/9781315568935`.

[391] Wright, O. and Wright, W., “The Wright Brothers’ Aeroplane,” *Aeronautical journal (London, England : 1897)*, vol. 20, no. 79, pp. 100–106, 1916, ISSN: 2398-1873. DOI: `10.1017/S2398187300142525`.

[392] Wrigley, S., *Hard Landing Within Limits*, 2020. [Online]. Available: `https://fearoflanding.com/accidents/accident-reports/hard-landing-within-limits/` (visited on 07/05/2022).

[393] Wynne, R. A., Beanland, V., and Salmon, P. M., "Systematic review of driving simulator validation studies," *Safety Science*, vol. 117, pp. 138–151, 2019, ISSN: 09257535. DOI: `10.1016/j.ssci.2019.04.004`.

[394] Xiong, J., Satoshi, M., and Fukumoto, K., "The Effects of Touch Button Size on Touchscreen Operability," *Journal of Mechanics Engineering and Automation*, vol. 4, no. 8, 2014, ISSN: 21595275. DOI: `10.17265/2159-5275/2014.08.006`.

[395] Zaychik, L. E., Grinev, K. N., Yashin, Y. P., and Sorokin, S. A., "Effect of Feel System Characteristics on Pilot Model Parameters," *IFAC-PapersOnLine*, vol. 49, no. 32, pp. 165–170, 2016, ISSN: 24058963. DOI: `10.1016/j.ifacol.2016.12.208`.

[396] Zazulia, N., *New Honeywell Touchscreen Will 'Delight' Gulfstream G500/600 Customers*, 2018. [Online]. Available: `https://www.aviationtoday.com/2018/07/27/honeywell-g500-symmetry/` (visited on 12/21/2022).

[397] Zhang, J., Roumeliotis, I., and Zolotas, A., "Model-based fully coupled propulsion-aerodynamics optimization for hybrid electric aircraft energy management strategy," *Energy*, vol. 245, p. 123 239, 2022, ISSN: 03605442. DOI: `10.1016/j.energy.2022.123239`.

[398] Zhang, Y., Chen, H., Li, W., Wang, Z., and Xin, X., "Research on the optimal control-display ratio of smart phone touch screen," in *Advances in Intelligent Systems and Computing*, vol. 777, 2019, pp. 382–386, ISBN: 9783319947051. DOI: `10.1007/978-3-319-94706-8_42`.

[399] Zhang, Z., "Microsoft Kinect Sensor and Its Effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012, ISSN: 1070-986X. DOI: `10.1109/MMUL.2012.24`.

[400] Zhao, W., Gupta, A., Regan, C. D., Miglani, J., Kapania, R. K., and Seiler, P. J., "Component data assisted finite element model updating of composite flying-wing aircraft using multi-level optimization," *Aerospace Science and Technology*, vol. 95, p. 105 486, 2019, ISSN: 12709638. DOI: `10.1016/j.ast.2019.105486`.

[401] Zheng, Y. and Lei, X. Y., "Research and implementation of virtual cockpit panel development platform based on ARINC 661," in *Proceedings of 2014 IEEE Chinese Guidance, Navigation and Control Conference*, IEEE, 2014, pp. 1357–1361, ISBN: 978-1-4799-4699-0. DOI: `10.1109/CGNCC.2014.7007394`.

[402] Zuccaro, J., "The Flight Simulator for Advanced Aircraft - A new aeronautical research tool," in *Visual and Motion Simualtion Technology Conference*, Reston, Virigina: American Institute of Aeronautics and Astronautics, 1970, pp. 1–5. DOI: `10.2514/6.1970-359`.

List of Figures

List of Tables

Appendix A

Supplementary results

Scenario correlations and hierarchical clustering

Table A.1: Results of Bivariate Pearson Correlation R with two-tailed test of significance p to investigate if the averaged results of different types of questionnaires across scenarios can be correlated. Column "Score type" refers to the averaged result (average of DRV, DRH, LN and LD), and column R shows the correlation. $R > .700$ shows that there is a strong correlation between the averaged result and the result of each scenario [51]. Inc. - Inceptor.

		DRV		DRH		LN		LD	
Inc.	Score type	R	p	R	p	R	p	R	p
SS	CHR	.874	<.001	.780	<.001	.894	<.001	.876	<.001
	SUS-U	.843	<.001	.717	<.001	.935	<.001	.911	<.001
	SUS-L	.880	<.001	.859	<.001	.932	<.001	.918	<.001
	SUS-Total	.842	<.001	.730	<.001	.931	<.001	.910	<.001
	SART-D	.817	<.001	.742	<.001	.784	<.001	.831	<.001
	SART-S	.793	<.001	.874	<.001	.876	<.001	.727	<.001
	SART-U	.853	<.001	.799	<.001	.862	<.001	.864	<.001
	SART-Total	.830	<.001	.736	<.001	.849	<.001	.820	<.001
	NASA-TLX	.759	<.001	.716	<.001	.815	<.001	.714	<.001

continued . . .

Table A.1: …continued

		DRV		DRH		LN		LD	
Inc.	Score type	R	p	R	p	R	p	R	p
GP	CHR	.793	<.001	.828	<.001	.919	<.001	.843	<.001
	SUS-U	.879	<.001	.863	<.001	.947	<.001	.893	<.001
	SUS-L	.943	<.001	.953	<.001	.963	<.001	.938	<.001
	SUS-Total	.891	<.001	.891	<.001	.952	<.001	.902	<.001
	SART-D	.808	<.001	.728	<.001	.806	<.001	.777	<.001
	SART-S	.910	<.001	.873	<.001	.877	<.001	.718	<.001
	SART-U	.924	<.001	.909	<.001	.926	<.001	.897	<.001
	SART-Total	.875	<.001	.839	<.001	.882	<.001	.762	<.001
	NASA-TLX	.820	<.001	.815	<.001	.802	<.001	.789	<.001
TS	CHR	.809	<.001	.794	<.001	.890	<.001	.852	<.001
	SUS-U	.860	<.001	.857	<.001	.927	<.001	.886	<.001
	SUS-L	.932	<.001	.899	<.001	.941	<.001	.922	<.001
	SUS-Total	.866	<.001	.867	<.001	.920	<.001	.881	<.001
	SART-D	.865	<.001	.786	<.001	.858	<.001	.748	<.001
	SART-S	.873	<.001	.788	<.001	.873	<.001	.817	<.001
	SART-U	.955	<.001	.914	<.001	.947	<.001	.945	<.001
	SART-Total	.914	<.001	.859	<.001	.921	<.001	.885	<.001
	NASA-TLX	.831	<.001	.783	<.001	.870	<.001	.806	<.001

Initial factor analyses

Table A.2: Results of two-way rANOVAs for a combination of each group and factor. Inc. - inceptor, M. p - Mauchly's p, df E - df Error.

Group	Factor	M. p	df	df E	F	p	η_p^2
Gender	CHR	.089	2	70	0.412	.664	.012
	SUS-U	.532	2	70	1.390	.256	.038
	SUS-L	.998	2	70	2.550	.085	.068
	SUS-Total	.604	2	70	2.134	.126	.057

continued …

Table A.2: . . . continued

Group	Factor	M. p	df	df E	F	p	η_p^2
Gender	SART-D	.356	2	70	0.208	.813	.006
	SART-S	.612	2	70	0.049	.952	.001
	SART-U	.479	2	70	1.392	.255	.038
	SART-Total	.307	2	70	0.873	.422	.024
	NASA-TLX	.618	2	69	1.398	.254	.039
	PS, DRV	.245	2	70	6.824	.002	.163
	PS, DRH	.281	2	70	2.297	.108	.062
	PS, LN	.229	2	70	2.098	.130	.057
	PS, LD	.081	2	70	5.388	.007	.133
Handedness	CHR	.069	2	70	0.202	.818	.006
	SUS-U	.554	2	70	0.701	.500	.020
	SUS-L	.915	2	70	0.206	.814	.006
	SUS-Total	.559	2	70	0.460	.633	.013
	SART-D	.258	2	70	0.196	.823	.006
	SART-S	.467	2	70	0.028	.972	.001
	SART-U	.205	2	70	0.426	.655	.012
	SART-Total	.217	2	70	0.247	.782	.007
	NASA-TLX	.547	2	69	0.110	.896	.003
	PS, DRV	.165	2	70	1.762	.179	.048
	PS, DRH	.452	2	70	0.170	.844	.005
	PS, LN	.288	2	70	2.737	.072	.073
	PS, LD	.038	2	70	1.286	.283	.035
TS att.	CHR	.098	1	70	0.293	.590	.004
	SUS-U	.453	1	70	4.362	.040	.059
	SUS-L	.995	1	70	0.010	.922	.000
	SUS-Total	.512	1	70	2.763	.101	.038
	SART-D	.229	1	70	0.066	.798	.001
	SART-S	.790	1	70	0.003	.955	.000
	SART-U	.329	1	70	0.621	.433	.009
	SART-Total	.421	1	70	0.110	.741	.002
	NASA-TLX	.644	1	70	0.841	.362	.012

continued . . .

Table A.2: . . . continued

Group	Factor	M. p	*df*	*df* E	F	p	η_p^2
TS att.	PS, DRV	.294	1	70	0.283	.596	.004
	PS, DRH	.295	1	70	0.496	.484	.007
	PS, LN	.446	1	70	0.184	.669	.003
	PS, LD	.107	1	70	0.635	.428	.009
VG freq.	CHR	.168	4	68	0.575	.682	.033
	SUS-U	.306	4	68	1.165	.334	.064
	SUS-L	.809	4	68	1.362	.256	.074
	SUS-Total	.296	4	68	1.291	.282	.071
	SART-D	.389	4	68	0.390	.815	.022
	SART-S	.682	4	68	0.815	.520	.046
	SART-U	.283	4	68	0.525	.718	.030
	SART-Total	.235	4	68	0.485	.746	.028
	NASA-TLX	.712	4	67	1.506	.210	.082
	PS, DRV	.413	4	68	2.467	.053	.127
	PS, DRH	.336	4	68	0.974	.428	.054
	PS, LN	.590	4	68	1.110	.359	.061
	PS, LD	.037	4	68	0.871	.486	.049
GP usage	CHR	.080	1	71	0.430	.514	.006
	SUS-U	.459	1	71	0.260	.612	.004
	SUS-L	.925	1	71	1.293	.259	.018
	SUS-Total	.462	1	71	0.597	.442	.008
	SART-D	.335	1	71	1.676	.200	.023
	SART-S	.690	1	71	0.790	.377	.011
	SART-U	.327	1	71	0.010	.921	.000
	SART-Total	.319	1	71	0.016	.901	.000
	NASA-TLX	.578	1	70	0.048	.828	.001
	PS, DRV	.230	1	71	0.401	.529	.006
	PS, DRH	.256	1	71	0.100	.752	.001
	PS, LN	.430	1	71	1.688	.198	.023
	PS, LD	.115	1	71	1.133	.291	.016

continued . . .

Table A.2: ...continued

Group	Factor	M. p	df	df E	F	p	η_p^2
MG usage	CHR	.113	2	70	0.050	.951	.001
	SUS-U	.164	2	70	0.141	.868	.004
	SUS-L	.172	2	70	1.820	.170	.049
	SUS-Total	.108	2	70	0.530	.591	.015
	SART-D	.474	2	70	1.751	.181	.048
	SART-S	.464	2	70	1.417	.249	.039
	SART-U	.238	2	70	0.688	.506	.019
	SART-Total	.185	2	70	0.404	.669	.011
	NASA-TLX	.590	2	69	0.344	.710	.010
	PS, DRV	.298	2	70	1.658	.198	.045
	PS, DRH	.454	2	70	0.709	.496	.020
	PS, LN	.352	2	70	2.121	.128	.057
	PS, LD	.020	2	70	2.274	.110	.061
Inc. order	CHR	.150	1	71	0.031	.860	.000
	SUS-U	.453	1	71	0.029	.864	.000
	SUS-L	.987	1	71	0.045	.833	.001
	SUS-Total	.570	1	71	0.042	.838	.001
	SART-D	.287	1	71	1.494	.226	.021
	SART-S	.719	1	71	0.029	.866	.000
	SART-U	.292	1	71	0.508	.478	.007
	SART-Total	.335	1	71	1.449	.233	.020
	NASA-TLX	.298	1	70	0.267	.607	.004
	PS, DRV	.131	1	71	1.709	.195	.024
	PS, DRH	.311	1	71	0.113	.738	.002
	PS, LN	.595	1	71	3.108	.082	.042
	PS, LD	.101	1	71	1.199	.277	.017
FG gr.	CHR	.319	2	70	1.587	.212	.043
	SUS-U	.454	2	70	5.822	.005	.143
	SUS-L	.818	2	70	0.802	.453	.022
	SUS-Total	.557	2	70	2.809	.067	.074
	SART-D	.540	2	70	0.602	.551	.017

continued ...

Table A.2: . . . continued

Group	Factor	M. p	df	df E	F	p	η_p^2
FG gr.	SART-S	.554	2	70	0.655	.523	.018
	SART-U	.239	2	70	0.190	.828	.005
	SART-Total	.124	2	70	0.602	.551	.017
	NASA-TLX	.353	2	69	0.765	.469	.022
	PS, DRV	.416	2	70	14.034	.000	.286
	PS, DRH	.353	2	70	4.095	.021	.105
	PS, LN	.562	2	70	5.695	.005	.140
	PS, LD	.106	2	70	10.191	.000	.226
VG gr.	CHR	.135	2	70	1.060	.352	.029
	SUS-U	.297	2	70	2.055	.136	.055
	SUS-L	.798	2	70	2.539	.086	.068
	SUS-Total	.274	2	70	2.554	.085	.068
	SART-D	.397	2	70	0.548	.581	.015
	SART-S	.647	2	70	0.811	.449	.023
	SART-U	.369	2	70	0.148	.863	.004
	SART-Total	.226	2	70	0.071	.932	.002
	NASA-TLX	.724	2	69	2.926	.060	.078
	PS, DRV	.315	2	70	0.427	.654	.012
	PS, DRH	.329	2	70	0.270	.764	.008
	PS, LN	.413	2	70	0.470	.627	.013
	PS, LD	.084	2	70	0.300	.741	.009

Age correlation

Table A.3: Results of Bivariate Pearson Correlation R with one-tailed test of significance p to investigate the correlation between participant's age and results of SART-Understanding and performance score (PS). Bold values indicate trend at level $p < .1$ or significance at level $p < .5$ and lower.

		avg		DRV		DRH		LN		LD	
Measure	Inc.	R	p	R	p	R	p	R	p	R	p
SART-U	SS	**-.160**	**.089**	**-.211**	**.037**	-.148	.108	-.107	.186	-.109	.182
	GP	**-.177**	**.066**	**-.244**	**.019**	**-.160**	**.090**	-.150	.105	-.101	.199
	TS	.036	.381	.018	.442	.050	.340	.033	.391	.084	.242
PS	SS	-	-	-.103	.193	-.109	.179	-.059	.310	.090	.225
	GP	-	-	-.057	.315	**-.248**	**.016**	.004	.485	**-.257**	**.014**
	TS	-	-	**-.361**	**.001**	**-.286**	**.007**	**-.297**	**.005**	**-.167**	**.078**

Table A.4: Results of Bivariate Pearson Correlation R with one-tailed test of significance p to further investigate the correlation between participant's age and results, based upon Tab. A.3. Here, CHR, SUS-U, SUS-L, SUS-Total, SART-D, SART-S, SART-Total, and NASA-TLX are investigated. Analyses of specific scenarios were only performed when the average value showed a trend in correlation, at least at elevated tolerance level of $p < .1$. Bold values indicate trend at level $p < .1$ or significance at level $p < .5$ and lower.

		avg		DRV		DRH		LN		LD	
Score type	Inc.	R	p	R	p	R	p	R	p	R	p
	SS	**.160**	**.089**	.114	.171	**.251**	**.017**	**.178**	**.067**	.085	.238
CHR	GP	**.234**	**.022**	**.234**	**.024**	**.202**	**.045**	**.191**	**.054**	.099	.204
	TS	.003	.491								
	SS	.024	.420								
SUS-U	GP	**-.285**	**.007**	**-.340**	**.002**	**-.373**	**.001**	**-.229**	**.027**	-.095	.214
	TS	.006	.481								
	SS	**.206**	**.041**	.080	.251	.097	.208	**.250**	**.017**	**.298**	**.005**
SUS-L	GP	**-.190**	**.052**	**-.324**	**.003**	**-.317**	**.003**	**-.176**	**.070**	-.086	.236
	TS	-.016	.445								
	SS	.080	.250								
SUS-Total	GP	**-.287**	**.007**	**-.367**	**.001**	**-.395**	$< .001$	**-.235**	**.023**	-.102	.198
	TS	$< .001$	.498								

continued ...

Table A.4: . . . continued

		avg		DRV		DRH		LN		LD	
Score type	Inc.	R	p	R	p	R	p	R	p	R	p
SART-D	SS	-.004	.486								
	GP	**.335**	**.002**	**.449**	$< .001$	**.363**	**.001**	**.242**	**.020**	.022	.428
	TS	-.066	.288								
SART-S	SS	.081	.249								
	GP	**.152**	**.098**	**.202**	**.045**	**.156**	**.096**	.075	.266	.040	.371
	TS	.130	.134								
SART-Total	SS	-.036	.382								
	GP	**-.229**	**.025**	**-.303**	**.005**	**-.211**	**.038**	**-.211**	**.037**	-.057	.317
	TS	.120	.154								
NASA-TLX	SS	**.246**	**.019**	**.197**	**.048**	**.306**	**.004**	**.172**	**.074**	.060	.308
	GP	**.346**	**.001**	**.367**	**.001**	**.298**	**.005**	**.182**	**.063**	**.257**	**.015**
	TS	.035	.385								

Table A.5: Results of Bivariate Spearman's Correlation ρ with one-tailed test of significance p to investigate the correlation between participant's age and results. Analyses of specific scenarios were only performed when the average value showed trending correlation, at least at an elevated tolerance level of $p < .1$. Bold values indicate trend at level $p < .1$ or significance at level $p < .5$ and lower.

		avg		DRV		DRH		LN		LD	
Score type	Inc.	ρ	p	ρ	p	ρ	p	ρ	p	ρ	p
CHR	SS	.151	.100								
	GP	.017	.442								
	TS	.001	.496								
SUS-U	SS	-.019	.436								
	GP	-.080	.249								
	TS	-.062	.300								
SUS-L	SS	**.234**	**.023**	.107	.186	.087	.232	**.263**	**.013**	**.336**	**.002**
	GP	-.014	.453								
	TS	-.012	.460								
SUS-Total	SS	.061	.305								
	GP	-.066	.289								
	TS	-.046	.348								
SART-D	SS	.086	.235								
	GP	**.273**	**.009**	**.386**	$< .001$	**.272**	**.010**	**.231**	**.026**	.014	.453
	TS	.020	.432								

continued ...

Table A.5: … continued

		avg		DRV		DRH		LN		LD	
Score type	Inc.	ρ	p	ρ	p	ρ	p	ρ	p	ρ	p
SART-S	SS	.109	.178								
	GP	**.164**	**.081**	**.186**	**.059**	**.164**	**.085**	**.173**	**.073**	-.004	.487
	TS	.111	.174								
SART-U	SS	-.081	.247								
	GP	-.008	.472								
	TS	.038	.375								
SART-Total	SS	.009	.470								
	GP	-.113	.168								
	TS	.069	.279								
NASA-TLX	SS	**.198**	**.048**	**.180**	**.065**	**.206**	**.042**	**.157**	**.094**	.064	.296
	GP	**.273**	**.010**	**.321**	**.003**	**.256**	**.015**	**.211**	**.038**	.120	.158
	TS	**.171**	**.076**	.065	.293	**.305**	**.005**	**.217**	**.034**	-.008	.474
PS	SS	-	-	**-.152**	**.099**	**-.194**	**.050**	**-.157**	**.092**	.061	.303
	GP	-	-	**-.159**	**.088**	**-.247**	**.017**	-.059	.308	**-.260**	**.013**
	TS	-	-	**-.368**	**.001**	**-.336**	**.002**	**-.319**	**.003**	-.108	.180

Main effects and interactions

SUS-U with TS attitude

Table A.6: Estimates for SUS-U scores with TS attitude as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
TS att.	1	43.98	3.182	37.63	50.33
	2	46.20	3.182	39.85	52.56
	3	44.28	2.250	39.79	48.77
	4	47.08	2.250	42.59	51.57
	5	51.35	2.250	46.86	55.85
Inc.	SS	56.97	1.596	53.78	60.15
	GP	52.51	1.979	48.56	56.46
	TS	30.27	2.090	26.10	34.44
1	* SS	54.93	4.265	46.42	63.44
	* GP	50.49	5.288	39.93	61.04
	* TS	26.53	5.586	15.38	37.68
2	* SS	55.69	4.265	47.18	64.21
	* GP	52.99	5.288	42.43	63.54
	* TS	29.93	5.586	18.78	41.08
3	* SS	54.93	3.016	48.91	60.95
	* GP	52.47	3.739	45.00	59.93
	* TS	25.45	3.950	17.57	33.34
4	* SS	58.92	3.016	52.90	64.94
	* GP	51.81	3.739	44.34	59.27
	* TS	30.52	3.950	22.64	38.40
5	* SS	60.35	3.016	54.33	66.37
	* GP	54.79	3.739	47.33	62.26

continued …

Table A.6: . . . continued

Effect		M	SE	LB	UB
5	* TS	38.92	3.950	31.04	46.81

Table A.7: Pairwise comparisons of main effects of inceptor and TS attitude, based on estimated marginal means for SUS-U scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
TS att.	1	2	-2.222	4.501	1.000	-15.287	10.843
		3	-.301	3.898	1.000	-11.616	11.014
		4	-3.102	3.898	1.000	-14.417	8.213
		5	-7.373	3.898	.629	-18.687	3.942
	2	1	2.222	4.501	1.000	-10.843	15.287
		3	1.921	3.898	1.000	-9.393	13.236
		4	-.880	3.898	1.000	-12.194	10.435
		5	-5.150	3.898	1.000	-16.465	6.164
	3	1	.301	3.898	1.000	-11.014	11.616
		2	-1.921	3.898	1.000	-13.236	9.393
		4	-2.801	3.182	1.000	-12.039	6.437
		5	-7.072	3.182	.297	-16.310	2.167
	4	1	3.102	3.898	1.000	-8.213	14.417
		2	.880	3.898	1.000	-10.435	12.194
		3	2.801	3.182	1.000	-6.437	12.039
		5	-4.271	3.182	1.000	-13.509	4.968
	5	1	7.373	3.898	.629	-3.942	18.687
		2	5.150	3.898	1.000	-6.164	16.465
		3	7.072	3.182	.297	-2.167	16.310
		4	4.271	3.182	1.000	-4.968	13.509
Inc.	SS	GP	4.458	2.429	.212	-1.505	10.422
		TS	26.694	2.743	.000	19.960	33.429

continued . . .

Table A.7: . . . continued

Factor			MD	SE	p	LB	UB
Inc.	GP	SS	-4.458	2.429	.212	-10.422	1.505
		TS	22.236	2.512	.000	16.068	28.404
	TS	SS	-26.694	2.743	.000	-33.429	-19.960
		GP	-22.236	2.512	.000	-28.404	-16.068

Table A.8: Pairwise comparisons based on estimated marginal means for SUS-U scores with TS attitude as a grouping factor showing patterns of interaction of inceptor and TS attitude. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	1	2	-.764	6.032	1.000	-18.273	16.746
		3	.000	5.224	1.000	-15.164	15.164
		4	-3.993	5.224	1.000	-19.157	11.171
		5	-5.417	5.224	1.000	-20.580	9.747
	2	1	.764	6.032	1.000	-16.746	18.273
		3	.764	5.224	1.000	-14.400	15.928
		4	-3.229	5.224	1.000	-18.393	11.935
		5	-4.653	5.224	1.000	-19.816	10.511
	3	1	.000	5.224	1.000	-15.164	15.164
		2	-.764	5.224	1.000	-15.928	14.400
		4	-3.993	4.265	1.000	-16.374	8.388
		5	-5.417	4.265	1.000	-17.798	6.964
	4	1	3.993	5.224	1.000	-11.171	19.157
		2	3.229	5.224	1.000	-11.935	18.393
		3	3.993	4.265	1.000	-8.388	16.374
		5	-1.424	4.265	1.000	-13.805	10.957
	5	1	5.417	5.224	1.000	-9.747	20.580
		2	4.653	5.224	1.000	-10.511	19.816

continued . . .

Table A.8: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	5	3	5.417	4.265	1.000	-6.964	17.798
		4	1.424	4.265	1.000	-10.957	13.805
GP	1	2	-2.500	7.479	1.000	-24.211	19.211
		3	-1.979	6.477	1.000	-20.781	16.823
		4	-1.319	6.477	1.000	-20.122	17.483
		5	-4.306	6.477	1.000	-23.108	14.497
	2	1	2.500	7.479	1.000	-19.211	24.211
		3	.521	6.477	1.000	-18.281	19.323
		4	1.181	6.477	1.000	-17.622	19.983
		5	-1.806	6.477	1.000	-20.608	16.997
	3	1	1.979	6.477	1.000	-16.823	20.781
		2	-.521	6.477	1.000	-19.323	18.281
		4	.660	5.288	1.000	-14.692	16.012
		5	-2.326	5.288	1.000	-17.678	13.026
	4	1	1.319	6.477	1.000	-17.483	20.122
		2	-1.181	6.477	1.000	-19.983	17.622
		3	-.660	5.288	1.000	-16.012	14.692
		5	-2.986	5.288	1.000	-18.338	12.366
	5	1	4.306	6.477	1.000	-14.497	23.108
		2	1.806	6.477	1.000	-16.997	20.608
		3	2.326	5.288	1.000	-13.026	17.678
		4	2.986	5.288	1.000	-12.366	18.338
TS	1	2	-3.403	7.900	1.000	-26.335	19.529
		3	1.076	6.841	1.000	-18.783	20.936
		4	-3.993	6.841	1.000	-23.853	15.867
		5	-12.396	6.841	.745	-32.256	7.464
	2	1	3.403	7.900	1.000	-19.529	26.335
		3	4.479	6.841	1.000	-15.381	24.339
		4	-.590	6.841	1.000	-20.450	19.269

continued . . .

Table A.8: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
TS	2	5	-8.993	6.841	1.000	-28.853	10.867
	3	1	-1.076	6.841	1.000	-20.936	18.783
		2	-4.479	6.841	1.000	-24.339	15.381
		4	-5.069	5.586	1.000	-21.285	11.146
		5	-13.472	5.586	.186	-29.688	2.743
	4	1	3.993	6.841	1.000	-15.867	23.853
		2	.590	6.841	1.000	-19.269	20.450
		3	5.069	5.586	1.000	-11.146	21.285
		5	-8.403	5.586	1.000	-24.618	7.813
	5	1	12.396	6.841	.745	-7.464	32.256
		2	8.993	6.841	1.000	-10.867	28.853
		3	13.472	5.586	.186	-2.743	29.688
		4	8.403	5.586	1.000	-7.813	24.618

Table A.9: Pairwise comparisons based on estimated marginal means for SUS-U scores with TS attitude as a grouping factor showing patterns of interaction of TS attitude and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
1	SS	GP	4.444	6.490	1.000	-11.493	20.382
		TS	28.403	7.330	.001	10.403	46.403
	GP	SS	-4.444	6.490	1.000	-20.382	11.493
		TS	23.958	6.713	.002	7.474	40.443
	TS	SS	-28.403	7.330	.001	-46.403	-10.403
		GP	-23.958	6.713	.002	-40.443	-7.474
2	SS	GP	2.708	6.490	1.000	-13.230	18.646
		TS	25.764	7.330	.002	7.764	43.764
	GP	SS	-2.708	6.490	1.000	-18.646	13.230

continued . . .

Table A.9: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
2	GP	TS	23.056	6.713	.003	6.571	39.540
	TS	SS	-25.764	7.330	.002	-43.764	-7.764
		GP	-23.056	6.713	.003	-39.540	-6.571
3	SS	GP	2.465	4.589	1.000	-8.805	13.735
		TS	29.479	5.183	.000	16.751	42.207
	GP	SS	-2.465	4.589	1.000	-13.735	8.805
		TS	27.014	4.747	.000	15.357	38.670
	TS	SS	-29.479	5.183	.000	-42.207	-16.751
		GP	-27.014	4.747	.000	-38.670	-15.357
4	SS	GP	7.118	4.589	.377	-4.152	18.388
		TS	28.403	5.183	.000	15.675	41.131
	GP	SS	-7.118	4.589	.377	-18.388	4.152
		TS	21.285	4.747	.000	9.628	32.941
	TS	SS	-28.403	5.183	.000	-41.131	-15.675
		GP	-21.285	4.747	.000	-32.941	-9.628
5	SS	GP	5.556	4.589	.691	-5.714	16.825
		TS	21.424	5.183	.000	8.696	34.151
	GP	SS	-5.556	4.589	.691	-16.825	5.714
		TS	15.868	4.747	.004	4.212	27.525
	TS	SS	-21.424	5.183	.000	-34.151	-8.696
		GP	-15.868	4.747	.004	-27.525	-4.212

Table A.10: Post-hoc pairwise comparisons based on observed means with different methods for SUS-U scores with TS attitude grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 91.149$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	1	2	-2.222	4.501	.988	-14.839	10.395
		3	-.301	3.898	1.000	-11.228	10.626

continued . . .

Table A.10: . . . continued

Method	Grouping factor		MD	SE	*p*	LB	UB
Tukey HSD	1	4	-3.102	3.898	.931	-14.029	7.825
		5	-7.373	3.898	.332	-18.300	3.554
	2	1	2.222	4.501	.988	-10.395	14.839
		3	1.921	3.898	.988	-9.006	12.848
		4	-.880	3.898	.999	-11.806	10.047
		5	-5.150	3.898	.679	-16.077	5.776
	3	1	.301	3.898	1.000	-10.626	11.228
		2	-1.921	3.898	.988	-12.848	9.006
		4	-2.801	3.182	.903	-11.723	6.121
		5	-7.072	3.182	.184	-15.993	1.850
	4	1	3.102	3.898	.931	-7.825	14.029
		2	.880	3.898	.999	-10.047	11.806
		3	2.801	3.182	.903	-6.121	11.723
		5	-4.271	3.182	.666	-13.193	4.651
	5	1	7.373	3.898	.332	-3.554	18.300
		2	5.150	3.898	.679	-5.776	16.077
		3	7.072	3.182	.184	-1.850	15.993
		4	4.271	3.182	.666	-4.651	13.193
Bonferroni	1	2	-2.222	4.501	1.000	-15.287	10.843
		3	-.301	3.898	1.000	-11.616	11.014
		4	-3.102	3.898	1.000	-14.417	8.213
		5	-7.373	3.898	.629	-18.687	3.942
	2	1	2.222	4.501	1.000	-10.843	15.287
		3	1.921	3.898	1.000	-9.393	13.236
		4	-.880	3.898	1.000	-12.194	10.435
		5	-5.150	3.898	1.000	-16.465	6.164
	3	1	.301	3.898	1.000	-11.014	11.616
		2	-1.921	3.898	1.000	-13.236	9.393
		4	-2.801	3.182	1.000	-12.039	6.437

continued . . .

Table A.10: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Bonferroni	3	5	-7.072	3.182	.297	-16.310	2.167
	4	1	3.102	3.898	1.000	-8.213	14.417
		2	.880	3.898	1.000	-10.435	12.194
		3	2.801	3.182	1.000	-6.437	12.039
		5	-4.271	3.182	1.000	-13.509	4.968
	5	1	7.373	3.898	.629	-3.942	18.687
		2	5.150	3.898	1.000	-6.164	16.465
		3	7.072	3.182	.297	-2.167	16.310
		4	4.271	3.182	1.000	-4.968	13.509

SUS-U with FE

Table A.11: Estimates for SUS-U scores with FE group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the FE group and the inceptor.

Effect		M	SE	LB	UB
FE gr.	A	40.57	2.126	36.33	44.81
	B	49.83	1.976	45.89	53.77
	C	48.18	1.638	44.92	51.45
Inceptor	SS	57.75	1.440	54.88	60.63
	GP	51.73	1.792	48.15	55.30
	TS	29.10	1.940	25.23	32.97
A *	SS	56.64	2.756	51.15	62.14
	GP	45.39	3.430	38.55	52.24
	TS	19.67	3.713	12.26	27.08
B *	SS	63.92	2.561	58.81	69.03
	GP	53.64	3.188	47.28	59.99

continued . . .

Table A.11: . . . continued

Effect		M	SE	LB	UB
B *	TS	31.93	3.451	25.05	38.81
C *	SS	52.70	2.123	48.46	56.93
	GP	56.15	2.643	50.88	61.42
	TS	35.70	2.861	30.00	41.41

Table A.12: Pairwise comparisons of main effects of inceptor and FE group (FE gr.) based on estimated marginal means for SUS-U scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
FE gr.	A	B	-9.259	2.902	.006	-16.378	-2.141
		C	-7.613	2.684	.018	-14.197	-1.030
	B	A	9.259	2.902	.006	2.141	16.378
		C	1.646	2.566	1.000	-4.649	7.941
	C	A	7.613	2.684	.018	1.030	14.197
		B	-1.646	2.566	1.000	-7.941	4.649
Inc.	SS	GP	6.026	2.133	.018	.793	11.258
		TS	28.652	2.379	$< .001$	22.817	34.486
	GP	SS	-6.026	2.133	.018	-11.258	-.793
		TS	22.626	2.416	$< .001$	16.699	28.552
	TS	SS	-28.652	2.379	$< .001$	-34.486	-22.817
		GP	-22.626	2.416	$< .001$	-28.552	-16.699

Table A.13: Pairwise comparisons based on estimated marginal means for SUS-U scores with FE group as a grouping factor showing patterns of interaction of Inceptor and FE group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	FE gr.		MD	SE	p	LB	UB
SS	A	B	-7.276	3.762	.171	-16.503	1.951
		C	3.949	3.479	.780	-4.584	12.482
	B	A	7.276	3.762	.171	-1.951	16.503
		C	11.225	3.327	.004	3.065	19.385
	C	A	-3.949	3.479	.780	-12.482	4.584
		B	-11.225	3.327	.004	-19.385	-3.065
GP	A	B	-8.242	4.683	.248	-19.727	3.244
		C	-10.758	4.330	.046	-21.379	-.136
	B	A	8.242	4.683	.248	-3.244	19.727
		C	-2.516	4.141	1.000	-12.673	7.641
	C	A	10.758	4.330	.046	.136	21.379
		B	2.516	4.141	1.000	-7.641	12.673
TS	A	B	-12.261	5.069	.055	-24.695	.174
		C	-16.032	4.688	.003	-27.531	-4.533
	B	A	12.261	5.069	.055	-.174	24.695
		C	-3.771	4.483	1.000	-14.767	7.225
	C	A	16.032	4.688	.003	4.533	27.531
		B	3.771	4.483	1.000	-7.225	14.767

Table A.14: Pairwise comparisons based on estimated marginal means for SUS-U scores with FE group as a grouping factor showing patterns of interaction of FE group and Inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

FE gr.	Inc.		MD	SE	p	LB	UB
A	SS	GP	11.250	4.082	.022	1.236	21.264

continued ...

Table A.14: . . . continued

FE gr.	Inc.		MD	SE	p	LB	UB
A	SS	TS	36.974	4.552	$< .001$	25.808	48.139
	GP	SS	-11.250	4.082	.022	-21.264	-1.236
		TS	25.724	4.624	$< .001$	14.382	37.066
	TS	SS	-36.974	4.552	$< .001$	-48.139	-25.808
		GP	-25.724	4.624	$< .001$	-37.066	-14.382
B	SS	GP	10.284	3.794	.025	.978	19.590
		TS	31.989	4.230	$< .001$	21.612	42.365
	GP	SS	-10.284	3.794	.025	-19.590	-.978
		TS	21.705	4.297	$< .001$	11.164	32.245
	TS	SS	-31.989	4.230	$< .001$	-42.365	-21.612
		GP	-21.705	4.297	$< .001$	-32.245	-11.164
C	SS	GP	-3.457	3.146	.827	-11.173	4.259
		TS	16.992	3.508	$< .001$	8.389	25.596
	GP	SS	3.457	3.146	.827	-4.259	11.173
		TS	20.449	3.563	$< .001$	11.710	29.189
	TS	SS	-16.992	3.508	$< .001$	-25.596	-8.389
		GP	-20.449	3.563	$< .001$	-29.189	-11.710

Table A.15: Post-hoc pairwise comparisons based on observed means with different methods for SUS-U scores with FE group (FE gr.). MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 85.871$.

Method	FE gr.		MD	SE	p	LB	UB
Tukey HSD	A	B	-9.259	2.902	.006	-16.209	-2.310
		C	-7.613	2.684	.016	-14.040	-1.187
	B	A	9.259	2.902	.006	2.310	16.209
		C	1.646	2.566	.798	-4.500	7.792
	C	A	7.613	2.684	.016	1.187	14.040
		B	-1.646	2.566	.798	-7.792	4.500

continued . . .

Table A.15: . . . continued

Method	FE gr.		MD	SE	p	LB	UB
Bonferroni	A	B	-9.259	2.902	.006	-16.378	-2.141
		C	-7.613	2.684	.018	-14.197	-1.030
	B	A	9.259	2.902	.006	2.141	16.378
		C	1.646	2.566	1.000	-4.649	7.941
	C	A	7.613	2.684	.018	1.030	14.197
		B	-1.646	2.566	1.000	-7.941	4.649

SUS-U with VG

Table A.16: Estimates for SUS-U scores with VG group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
VG gr.	A	47.28	1.644	44.00	50.56
	B	48.53	1.985	44.57	52.49
	C	42.10	2.599	36.91	47.28
Inceptor	SS	56.69	1.618	53.46	59.92
	GP	50.90	1.825	47.26	54.54
	TS	30.32	2.198	25.94	34.71
A *	SS	57.55	2.179	53.21	61.90
	GP	53.75	2.459	48.85	58.65
	TS	30.54	2.961	24.63	36.44
B *	SS	57.92	2.632	52.67	63.17
	GP	57.29	2.969	51.37	63.21
	TS	30.39	3.575	23.26	37.52
C *	SS	54.60	3.446	47.73	61.47
	GP	41.65	3.887	33.90	49.40
	TS	30.04	4.681	20.71	39.38

Table A.17: Pairwise comparisons of main effects of inceptor and VG group based on estimated marginal means for SUS-U scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Effect			MD	SE	p	LB	UB
VG gr.	A	B	-1.253	2.578	1.000	-7.576	5.070
		C	5.182	3.076	.290	-2.363	12.726
	B	A	1.253	2.578	1.000	-5.070	7.576
		C	6.435	3.271	.159	-1.588	14.458
	C	A	-5.182	3.076	.290	-12.726	2.363
		B	-6.435	3.271	.159	-14.458	1.588
Inc.	SS	GP	5.792	2.343	.048	.044	11.539
		TS	26.366	2.729	$< .001$	19.673	33.059
	GP	SS	-5.792	2.343	.048	-11.539	-.044
		TS	20.574	2.447	$< .001$	14.571	26.578
	TS	SS	-26.366	2.729	$< .001$	-33.059	-19.673
		GP	-20.574	2.447	$< .001$	-26.578	-14.571

Table A.18: Pairwise comparisons based on estimated marginal means for SUS-U scores with VG group as a grouping factor showing patterns of interaction of inceptor and VG group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	FE gr.		MD	SE	p	LB	UB
SS	A	B	-.363	3.417	1.000	-8.745	8.018
		C	2.955	4.077	1.000	-7.045	12.956
	B	A	.363	3.417	1.000	-8.018	8.745
		C	3.318	4.336	1.000	-7.317	13.954
	C	A	-2.955	4.077	1.000	-12.956	7.045
		B	-3.318	4.336	1.000	-13.954	7.317

continued . . .

Table A.18: . . . continued

Inc.	FE gr.		MD	SE	p	LB	UB
GP	A	B	-3.542	3.855	1.000	-12.997	5.914
		C	12.098	4.600	.031	.816	23.380
	B	A	3.542	3.855	1.000	-5.914	12.997
		C	15.640	4.891	.006	3.642	27.638
	C	A	-12.098	4.600	.031	-23.380	-.816
		B	-15.640	4.891	.006	-27.638	-3.642
TS	A	B	.145	4.642	1.000	-11.241	11.532
		C	.491	5.539	1.000	-13.095	14.077
	B	A	-.145	4.642	1.000	-11.532	11.241
		C	.346	5.890	1.000	-14.103	14.795
	C	A	-.491	5.539	1.000	-14.077	13.095
		B	-.346	5.890	1.000	-14.795	14.103

Table A.19: Pairwise comparisons based on estimated marginal means for SUS-U scores with VG group as a grouping factor showing patterns of interaction of VG group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

FE gr.	Inc.		MD	SE	p	LB	UB
A	SS	GP	3.804	3.157	.697	-3.940	11.547
		TS	27.018	3.676	$< .001$	18.001	36.035
	GP	SS	-3.804	3.157	.697	-11.547	3.940
		TS	23.214	3.297	$< .001$	15.126	31.302
	TS	SS	-27.018	3.676	$< .001$	-36.035	-18.001
		GP	-23.214	3.297	$< .001$	-31.302	-15.126
B	SS	GP	.625	3.812	1.000	-8.726	9.976
		TS	27.526	4.439	$< .001$	16.637	38.415
	GP	SS	-.625	3.812	1.000	-9.976	8.726
		TS	26.901	3.982	$< .001$	17.134	36.668
	TS	SS	-27.526	4.439	$< .001$	-38.415	-16.637

continued . . .

Table A.19: . . . continued

FE gr.	Inc.		MD	SE	p	LB	UB
B	TS	GP	-26.901	3.982	< .001	-36.668	-17.134
C	SS	GP	12.946	4.992	.035	.703	25.190
		TS	24.554	5.813	< .001	10.296	38.811
	GP	SS	-12.946	4.992	.035	-25.190	-.703
		TS	11.607	5.214	.088	-1.181	24.396
	TS	SS	-24.554	5.813	< .001	-38.811	-10.296
		GP	-11.607	5.214	.088	-24.396	1.181

Table A.20: Post-hoc pairwise comparisons based on observed means with different methods for SUS-U scores with VG group. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 94.599$.

Method	VG gr.		MD	SE	p	LB	UB
Tukey HSD	A	B	-1.253	2.578	.878	-7.426	4.919
		C	5.182	3.076	.218	-2.183	12.547
	B	A	1.253	2.578	.878	-4.919	7.426
		C	6.435	3.271	.128	-1.398	14.267
	C	A	-5.182	3.076	.218	-12.547	2.183
		B	-6.435	3.271	.128	-14.267	1.398
Bonferroni	A	B	-1.253	2.578	1.000	-7.576	5.070
		C	5.182	3.076	.290	-2.363	12.726
	B	A	1.253	2.578	1.000	-5.070	7.576
		C	6.435	3.271	.159	-1.588	14.458
	C	A	-5.182	3.076	.290	-12.726	2.363
		B	-6.435	3.271	.159	-14.458	1.588

SUS-L with gender

Table A.21: Estimates for SUS-L scores with gender as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Gender	Female	11.98	.925	10.13	13.82
	Male	13.52	.534	12.46	14.59
	Pref. not to say	20.00	3.926	12.17	27.83
Inceptor	SS	15.57	1.564	12.45	18.69
	GP	16.42	1.737	12.95	19.88
	TS	13.51	1.868	9.79	17.24
Female *	SS	11.98	1.067	9.85	14.11
	GP	13.92	1.185	11.56	16.29
	TS	10.03	1.275	7.49	12.58
Male *	SS	14.73	.616	13.50	15.96
	GP	15.32	.684	13.96	16.69
	TS	10.51	.736	9.04	11.98
Pref. not to say *	SS	20.00	4.528	10.97	29.03
	GP	20.00	5.027	9.97	30.03
	TS	20.00	5.408	9.21	30.79

Table A.22: Pairwise comparisons of main effects of inceptor and gender, based on estimated marginal means for SUS-L scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Effect			MD	SE	p	LB	UB
Gender	Female	Male	-1.543	1.069	.459	-4.164	1.078
		Pref. not to say	-8.021	4.034	.152	-17.916	1.874
	Male	Female	1.543	1.069	.459	-1.078	4.164
		Pref. not to say	-6.478	3.963	.320	-16.197	3.242
	Pref. not to say	Female	8.021	4.034	.152	-1.874	17.916

continued ...

Table A.22: . . . continued

Effect			MD	SE	p	LB	UB
Gender	Pref. not to say	Male	6.478	3.963	.320	-3.242	16.197
Inc.	SS	GP	-.845	1.848	1.000	-5.378	3.688
		TS	2.056	1.860	.818	-2.505	6.618
	GP	SS	.845	1.848	1.000	-3.688	5.378
		TS	2.901	1.851	.365	-1.640	7.442
	TS	SS	-2.056	1.860	.818	-6.618	2.505
		GP	-2.901	1.851	.365	-7.442	1.640

Table A.23: Pairwise comparisons based on estimated marginal means for SUS-L scores with gender as a grouping factor showing patterns of interaction of inceptor and gender. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Gender		MD	SE	p	LB	UB
SS	Female	Male	-2.755	1.232	.086	-5.777	.268
		Pref. not to say	-8.021	4.652	.267	-19.431	3.390
	Male	Female	2.755	1.232	.086	-.268	5.777
		Pref. not to say	-5.266	4.569	.759	-16.475	5.942
	Pref. not to say	Female	8.021	4.652	.267	-3.390	19.431
		Male	5.266	4.569	.759	-5.942	16.475
GP	Female	Male	-1.400	1.368	.929	-4.757	1.956
		Pref. not to say	-6.076	5.165	.730	-18.746	6.593
	Male	Female	1.400	1.368	.929	-1.956	4.757
		Pref. not to say	-4.676	5.074	1.000	-17.121	7.769
	Pref. not to say	Female	6.076	5.165	.730	-6.593	18.746
		Male	4.676	5.074	1.000	-7.769	17.121
TS	Female	Male	-.475	1.472	1.000	-4.085	3.136
		Pref. not to say	-9.965	5.556	.232	-23.594	3.664
	Male	Female	.475	1.472	1.000	-3.136	4.085
		Pref. not to say	-9.491	5.458	.259	-22.878	3.897

continued . . .

Table A.23: . . . continued

Inc.	Gender		MD	SE	p	LB	UB
TS	Pref. not to say	Female	9.965	5.556	.232	-3.664	23.594
		Male	9.491	5.458	.259	-3.897	22.878

Table A.24: Pairwise comparisons based on estimated marginal means for SUS-L scores with gender as a grouping factor showing patterns of interaction of gender and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Gender	Inc.		MD	SE	p	LB	UB
Female	SS	GP	-1.944	1.261	.383	-5.037	1.148
		TS	1.944	1.269	.390	-1.168	5.057
	GP	SS	1.944	1.261	.383	-1.148	5.037
		TS	3.889	1.263	.009	.791	6.987
	TS	SS	-1.944	1.269	.390	-5.057	1.168
		GP	-3.889	1.263	.009	-6.987	-.791
Male	SS	GP	-.590	.728	1.000	-2.376	1.195
		TS	4.225	.733	< .001	2.428	6.022
	GP	SS	.590	.728	1.000	-1.195	2.376
		TS	4.815	.729	< .001	3.026	6.604
	TS	SS	-4.225	.733	< .001	-6.022	-2.428
		GP	-4.815	.729	< .001	-6.604	-3.026
Pref. not to say	SS	GP	< .001	5.350	1.000	-13.122	13.122
		TS	< .001	5.384	1.000	-13.205	13.205
	GP	SS	< .001	5.350	1.000	-13.122	13.122
		TS	< .001	5.359	1.000	-13.145	13.145
	TS	SS	< .001	5.384	1.000	-13.205	13.205
		GP	< .001	5.359	1.000	-13.145	13.145

SUS-L with MG

Table A.25: Estimates for SUS-L scores with MG usage as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
MG	no / h. ever	14.03	.736	12.56	15.50
	used to	11.85	.887	10.09	13.62
	yes	13.41	.809	11.80	15.03
Inceptor	SS	13.83	.506	12.82	14.84
	GP	15.11	.582	13.95	16.27
	TS	10.36	.641	9.08	11.64
no / h. ever *	SS	16.51	.793	14.93	18.09
	GP	13.79	.912	11.97	15.61
	TS	11.79	1.006	9.78	13.79
used to *	SS	11.78	.955	9.88	13.69
	GP	14.63	1.098	12.43	16.82
	TS	9.16	1.211	6.74	11.57
yes *	SS	13.20	.872	11.46	14.94
	GP	16.90	1.003	14.90	18.90
	TS	10.13	1.106	7.92	12.34

Table A.26: Pairwise comparisons of main effects of inceptor and MG usage, based on estimated marginal means for SUS-L scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Effect			MD	SE	*p*	LB	UB
MG	no / h. ever	used to	2.176	1.152	.189	-.651	5.003
		yes	.619	1.094	1.000	-2.065	3.303
	used to	no / h. ever	-2.176	1.152	.189	-5.003	.651
		yes	-1.557	1.200	.596	-4.502	1.387
	yes	no / h. ever	-.619	1.094	1.000	-3.303	2.065

continued ...

Table A.26: . . . continued

Effect			MD	SE	p	LB	UB
MG	yes	used to	1.557	1.200	.596	-1.387	4.502
Inc.	SS	GP	-1.275	.527	.054	-2.568	.017
		TS	3.473	.641	$< .001$	1.901	5.044
	GP	SS	1.275	.527	.054	-.017	2.568
		TS	4.748	.587	$< .001$	3.309	6.187
	TS	SS	-3.473	.641	$< .001$	-5.044	-1.901
		GP	-4.748	.587	$< .001$	-6.187	-3.309

Table A.27: Pairwise comparisons based on estimated marginal means for SUS-L scores with MG usage as a grouping factor showing patterns of interaction of inceptor and MG usage. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	MG		MD	SE	p	LB	UB
SS	no / h. ever	used to	4.727	1.242	.001	1.682	7.773
		yes	3.305	1.179	.020	.414	6.197
	used to	no / h. ever	-4.727	1.242	.001	-7.773	-1.682
		yes	-1.422	1.293	.826	-4.594	1.751
	yes	no / h. ever	-3.305	1.179	.020	-6.197	-.414
		used to	1.422	1.293	.826	-1.751	4.594
GP	no / h. ever	used to	-.832	1.428	1.000	-4.334	2.670
		yes	-3.108	1.355	.075	-6.433	.217
	used to	no / h. ever	.832	1.428	1.000	-2.670	4.334
		yes	-2.276	1.487	.391	-5.924	1.372
	yes	no / h. ever	3.108	1.355	.075	-.217	6.433
		used to	2.276	1.487	.391	-1.372	5.924
TS	no / h. ever	used to	2.633	1.574	.297	-1.229	6.494
		yes	1.659	1.495	.813	-2.008	5.325
	used to	no / h. ever	-2.633	1.574	.297	-6.494	1.229
		yes	-.974	1.640	1.000	-4.997	3.049

continued . . .

Table A.27: . . . continued

Inc.		MG	MD	SE	p	LB	UB
TS	yes	no / h. ever	-1.659	1.495	.813	-5.325	2.008
		used to	.974	1.640	1.000	-3.049	4.997

Table A.28: Pairwise comparisons based on estimated marginal means for SUS-L scores with MG usage as a grouping factor showing patterns of interaction of MG usage and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Gender	Inc.		MD	SE	p	LB	UB
Female	SS	GP	2.716	.826	.005	.688	4.743
		TS	4.720	1.005	$< .001$	2.255	7.185
	GP	SS	-2.716	.826	.005	-4.743	-.688
		TS	2.004	.920	.098	-.253	4.262
	TS	SS	-4.720	1.005	$< .001$	-7.185	-2.255
		GP	-2.004	.920	.098	-4.262	.253
Male	SS	GP	-2.844	.995	.017	-5.285	-.403
		TS	2.625	1.210	.100	-.343	5.593
	GP	SS	2.844	.995	.017	.403	5.285
		TS	5.469	1.108	$< .001$	2.750	8.187
	TS	SS	-2.625	1.210	.100	-5.593	.343
		GP	-5.469	1.108	$< .001$	-8.187	-2.750
Pref. not to say	SS	GP	-3.698	.908	$< .001$	-5.926	-1.470
		TS	3.073	1.105	.021	.363	5.783
	GP	SS	3.698	.908	$< .001$	1.470	5.926
		TS	6.771	1.012	$< .001$	4.289	9.252
	TS	SS	-3.073	1.105	.021	-5.783	-.363
		GP	-6.771	1.012	$< .001$	-9.252	-4.289

SUS-L with VG

Table A.29: Estimates for SUS-L scores with VG group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
	A	14.05	.664	12.72	15.37
VG gr.	B	13.19	.802	11.60	14.79
	C	11.25	1.050	9.16	13.34
	SS	14.01	.590	12.84	15.19
Inc.	GP	14.38	.596	13.20	15.57
	TS	10.09	.679	8.74	11.45
	SS	14.68	.795	13.09	16.26
A *	GP	15.95	.802	14.35	17.55
	TS	11.52	.915	9.69	13.34
	SS	13.44	.960	11.52	15.35
B *	GP	15.91	.969	13.98	17.84
	TS	10.23	1.104	8.03	12.44
	SS	13.93	1.258	11.42	16.44
C *	GP	11.29	1.269	8.76	13.82
	TS	8.53	1.446	5.64	11.41

Table A.30: Pairwise comparisons of main effects of inceptor and VG group, based on estimated marginal means for SUS-L scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Effect			MD	SE	*p*	LB	UB
VG gr.	A	B	.853	1.041	1.000	-1.700	3.406
		C	2.798	1.242	.082	-.248	5.844
	B	A	-.853	1.041	1.000	-3.406	1.700
		C	1.944	1.321	.436	-1.295	5.184
	C	A	-2.798	1.242	.082	-5.844	.248

continued . . .

Table A.30: . . . continued

Effect			MD	SE	p	LB	UB
VG gr.	C	B	-1.944	1.321	.436	-5.184	1.295
Inc.	SS	GP	-.369	.634	1.000	-1.925	1.187
		TS	3.922	.680	$< .001$	2.254	5.590
	GP	SS	.369	.634	1.000	-1.187	1.925
		TS	4.291	.665	$< .001$	2.660	5.923
	TS	SS	-3.922	.680	$< .001$	-5.590	-2.254
		GP	-4.291	.665	$< .001$	-5.923	-2.660

Table A.31: Pairwise comparisons based on estimated marginal means for SUS-L scores with VG group as a grouping factor showing patterns of interaction of inceptor and VG group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	VG gr.		MD	SE	p	LB	UB
SS	A	B	1.241	1.247	.969	-1.818	4.300
		C	.750	1.488	1.000	-2.900	4.400
	B	A	-1.241	1.247	.969	-4.300	1.818
		C	-.491	1.582	1.000	-4.373	3.390
	C	A	-.750	1.488	1.000	-4.400	2.900
		B	.491	1.582	1.000	-3.390	4.373
GP	A	B	.035	1.258	1.000	-3.051	3.121
		C	4.652	1.501	.008	.970	8.334
	B	A	-.035	1.258	1.000	-3.121	3.051
		C	4.617	1.596	.015	.701	8.532
	C	A	-4.652	1.501	.008	-8.334	-.970
		B	-4.617	1.596	.015	-8.532	-.701
TS	A	B	1.283	1.434	1.000	-2.234	4.801
		C	2.991	1.711	.254	-1.206	7.188
	B	A	-1.283	1.434	1.000	-4.801	2.234
		C	1.708	1.819	1.000	-2.755	6.170

continued . . .

Table A.31: . . . continued

Inc.	VG gr.		MD	SE	p	LB	UB
TS	C	A	-2.991	1.711	.254	-7.188	1.206
		B	-1.708	1.819	1.000	-6.170	2.755

Table A.32: Pairwise comparisons based on estimated marginal means for SUS-L scores with VG group as a grouping factor showing patterns of interaction of VG group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

VG gr.	Inc.		MD	SE	p	LB	UB
A	SS	GP	-1.268	.855	.427	-3.364	.829
		TS	3.161	.916	.003	.914	5.408
	GP	SS	1.268	.855	.427	-.829	3.364
		TS	4.429	.896	$< .001$	2.231	6.627
	TS	SS	-3.161	.916	.003	-5.408	-.914
		GP	-4.429	.896	$< .001$	-6.627	-2.231
B	SS	GP	-2.474	1.032	.058	-5.006	.058
		TS	3.203	1.106	.015	.490	5.917
	GP	SS	2.474	1.032	.058	-.058	5.006
		TS	5.677	1.082	$< .001$	3.023	8.331
	TS	SS	-3.203	1.106	.015	-5.917	-.490
		GP	-5.677	1.082	$< .001$	-8.331	-3.023
C	SS	GP	2.634	1.351	.166	-.681	5.949
		TS	5.402	1.448	.001	1.849	8.955
	GP	SS	-2.634	1.351	.166	-5.949	.681
		TS	2.768	1.417	.164	-.707	6.243
	TS	SS	-5.402	1.448	.001	-8.955	-1.849
		GP	-2.768	1.417	.164	-6.243	.707

Table A.33: Post-hoc pairwise comparisons based on observed means with different methods for SUS-L scores with VG group. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 15.420$.

Method	VG gr.		MD	SE	p	LB	UB
Tukey HSD	A	B	.853	1.041	.692	-1.639	3.345
		C	2.798	1.242	.069	-.176	5.771
	B	A	-.853	1.041	.692	-3.345	1.639
		C	1.944	1.321	.310	-1.218	5.107
	C	A	-2.798	1.242	.069	-5.771	.176
		B	-1.944	1.321	.310	-5.107	1.218
Bonferroni	A	B	.853	1.041	1.000	-1.700	3.406
		C	2.798	1.242	.082	-.248	5.844
	B	A	-.853	1.041	1.000	-3.406	1.700
		C	1.944	1.321	.436	-1.295	5.184
	C	A	-2.798	1.242	.082	-5.844	.248
		B	-1.944	1.321	.436	-5.184	1.295

SUS-Total with gender

Table A.34: Estimates for SUS-Total scores with gender as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Gender	Female	56.57	2.876	50.84	62.31
	Male	60.68	1.661	57.36	63.99
	Pref. not to say	80.00	12.204	55.66	104.34
Inc.	SS	77.65	5.227	67.23	88.08
	GP	75.25	6.466	62.36	88.15
	TS	44.34	7.076	30.23	58.45

continued …

Table A.34: . . . continued

Effect		M	SE	LB	UB
Female *	SS	63.30	3.566	56.19	70.41
	GP	65.31	4.411	56.51	74.11
	TS	41.11	4.828	31.48	50.74
Male *	SS	73.41	2.059	69.31	77.52
	GP	67.95	2.547	62.87	73.03
	TS	40.66	2.787	35.10	46.22
Pref. not to say *	SS	96.25	15.131	66.07	126.43
	GP	92.50	18.716	55.17	129.83
	TS	51.25	20.482	10.40	92.10

Table A.35: Pairwise comparisons of main effects of inceptor and gender, based on estimated marginal means for SUS-Total scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Gender	Female	Male	-4.101	3.321	.663	-12.248	4.046
		Pref. not to say	-23.426	12.538	.198	-54.180	7.329
	Male	Female	4.101	3.321	.663	-4.046	12.248
		Pref. not to say	-19.325	12.316	.363	-49.535	10.885
	Pref. not to say	Female	23.426	12.538	.198	-7.329	54.180
		Male	19.325	12.316	.363	-10.885	49.535
Inc.	SS	GP	2.400	7.756	1.000	-16.626	21.425
		TS	33.314	8.573	.001	12.286	54.342
	GP	SS	-2.400	7.756	1.000	-21.425	16.626
		TS	30.914	7.997	.001	11.298	50.531
	TS	SS	-33.314	8.573	.001	-54.342	-12.286
		GP	-30.914	7.997	.001	-50.531	-11.298

Table A.36: Pairwise comparisons based on estimated marginal means for SUS-Total scores with gender as a grouping factor showing patterns of interaction of inceptor and gender. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Gender		MD	SE	p	LB	UB
SS	Female	Male	-10.116	4.118	.050	-20.217	-.015
		Pref. not to say	-32.951	15.545	.113	-71.082	5.179
	Male	Female	10.116	4.118	.050	.015	20.217
		Pref. not to say	-22.836	15.270	.418	-60.291	14.620
	Pref. not to say	Female	32.951	15.545	.113	-5.179	71.082
		Male	22.836	15.270	.418	-14.620	60.291
GP	Female	Male	-2.639	5.094	1.000	-15.133	9.856
		Pref. not to say	-27.187	19.229	.485	-74.353	19.978
	Male	Female	2.639	5.094	1.000	-9.856	15.133
		Pref. not to say	-24.549	18.888	.594	-70.880	21.782
	Pref. not to say	Female	27.187	19.229	.485	-19.978	74.353
		Male	24.549	18.888	.594	-21.782	70.880
TS	Female	Male	.451	5.575	1.000	-13.222	14.125
		Pref. not to say	-10.139	21.044	1.000	-61.756	41.479
	Male	Female	-.451	5.575	1.000	-14.125	13.222
		Pref. not to say	-10.590	20.671	1.000	-61.294	40.114
	Pref. not to say	Female	10.139	21.044	1.000	-41.479	61.756
		Male	10.590	20.671	1.000	-40.114	61.294

Table A.37: Pairwise comparisons based on estimated marginal means for SUS-Total scores with gender as a grouping factor showing patterns of interaction of gender and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Gender	Inc.		MD	SE	p	LB	UB
Female	SS	GP	-2.014	5.292	1.000	-14.995	10.967

continued . . .

Table A.37: . . . continued

Gender	Inc.		MD	SE	p	LB	UB
Female	SS	TS	22.187	5.849	.001	7.840	36.535
	GP	SS	2.014	5.292	1.000	-10.967	14.995
		TS	24.201	5.456	< .001	10.817	37.585
	TS	SS	-22.187	5.849	.001	-36.535	-7.840
		GP	-24.201	5.456	< .001	-37.585	-10.817
Male	SS	GP	5.463	3.055	.234	-2.032	12.957
		TS	32.755	3.377	< .001	24.471	41.038
	GP	SS	-5.463	3.055	.234	-12.957	2.032
		TS	27.292	3.150	< .001	19.564	35.019
	TS	SS	-32.755	3.377	< .001	-41.038	-24.471
		GP	-27.292	3.150	< .001	-35.019	-19.564
Pref. not to say	SS	GP	3.750	22.453	1.000	-51.323	58.823
		TS	45.000	24.816	.222	-15.871	105.871
	GP	SS	-3.750	22.453	1.000	-58.823	51.323
		TS	41.250	23.150	.237	-15.533	98.033
	TS	SS	-45.000	24.816	.222	-105.871	15.871
		GP	-41.250	23.150	.237	-98.033	15.533

SUS-Total with TS attitude

Table A.38: Estimates for SUS-Total scores with TS attitude as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
TS att.	1	55.86	4.069	47.73	63.98
	2	61.85	4.069	53.73	69.97
	3	56.79	2.877	51.05	62.54
	4	60.58	2.877	54.84	66.32

continued . . .

Table A.38: . . . continued

Effect		M	SE	LB	UB
TS att.	5	64.50	2.877	58.76	70.25
Inc.	SS	71.08	1.989	67.11	75.05
	GP	67.51	2.404	62.71	72.31
	TS	41.15	2.484	36.19	46.11
TS att. 1 *	SS	67.71	5.316	57.10	78.32
	GP	63.61	6.426	50.79	76.44
	TS	36.25	6.639	23.00	49.50
TS att. 2 *	SS	71.46	5.316	60.85	82.07
	GP	69.31	6.426	56.48	82.13
	TS	44.79	6.639	31.54	58.04
TS att. 3 *	SS	67.47	3.759	59.96	74.97
	GP	67.64	4.544	58.57	76.71
	TS	35.28	4.695	25.91	44.65
TS att. 4 *	SS	73.89	3.759	66.39	81.39
	GP	67.43	4.544	58.36	76.50
	TS	40.42	4.695	31.05	49.79
TS att. 5 *	SS	74.90	3.759	67.39	82.40
	GP	69.58	4.544	60.51	78.65
	TS	49.03	4.695	39.66	58.40

Table A.39: Pairwise comparisons of main effects of TS attitude, based on estimated marginal means for SUS-Total scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

TS att.		MD	SE	p	LB	UB
1	2	-5.995	5.754	1.000	-22.700	10.710
	3	-.938	4.983	1.000	-15.404	13.529
	4	-4.722	4.983	1.000	-19.189	9.745

continued . . .

Table A.39: ...continued

	TS att.	MD	SE	p	LB	UB
1	5	-8.646	4.983	.874	-23.113	5.821
2	1	5.995	5.754	1.000	-10.710	22.700
	3	5.058	4.983	1.000	-9.409	19.525
	4	1.273	4.983	1.000	-13.194	15.740
	5	-2.650	4.983	1.000	-17.117	11.816
3	1	.938	4.983	1.000	-13.529	15.404
	2	-5.058	4.983	1.000	-19.525	9.409
	4	-3.785	4.069	1.000	-15.597	8.027
	5	-7.708	4.069	.625	-19.520	4.104
4	1	4.722	4.983	1.000	-9.745	19.189
	2	-1.273	4.983	1.000	-15.740	13.194
	3	3.785	4.069	1.000	-8.027	15.597
	5	-3.924	4.069	1.000	-15.736	7.889
5	1	8.646	4.983	.874	-5.821	23.113
	2	2.650	4.983	1.000	-11.816	17.117
	3	7.708	4.069	.625	-4.104	19.520
	4	3.924	4.069	1.000	-7.889	15.736

Table A.40: Pairwise comparisons of main effects of inceptor, based on estimated marginal means for SUS-Total scores with TS attitude as a grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

	Inc.	MD	SE	p	LB	UB
SS	GP	3.569	2.872	.655	-3.483	10.622
	TS	29.931	3.195	$< .001$	22.086	37.775
GP	SS	-3.569	2.872	.655	-10.622	3.483
	TS	26.361	2.900	$< .001$	19.241	33.482
TS	SS	-29.931	3.195	$< .001$	-37.775	-22.086
	GP	-26.361	2.900	$< .001$	-33.482	-19.241

Table A.41: Pairwise comparisons based on estimated marginal means for SUS-Total scores with TS attitude as a grouping factor showing patterns of interaction of inceptor and TS attitude. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	TS att.		MD	SE	p	LB	UB
SS	1	2	-3.750	7.518	1.000	-25.575	18.075
		3	.243	6.511	1.000	-18.658	19.144
		4	-6.181	6.511	1.000	-25.082	12.720
		5	-7.188	6.511	1.000	-26.088	11.713
	2	1	3.750	7.518	1.000	-18.075	25.575
		3	3.993	6.511	1.000	-14.908	22.894
		4	-2.431	6.511	1.000	-21.332	16.470
		5	-3.438	6.511	1.000	-22.338	15.463
	3	1	-.243	6.511	1.000	-19.144	18.658
		2	-3.993	6.511	1.000	-22.894	14.908
		4	-6.424	5.316	1.000	-21.856	9.009
		5	-7.431	5.316	1.000	-22.863	8.002
	4	1	6.181	6.511	1.000	-12.720	25.082
		2	2.431	6.511	1.000	-16.470	21.332
		3	6.424	5.316	1.000	-9.009	21.856
		5	-1.007	5.316	1.000	-16.440	14.426
	5	1	7.188	6.511	1.000	-11.713	26.088
		2	3.438	6.511	1.000	-15.463	22.338
		3	7.431	5.316	1.000	-8.002	22.863
		4	1.007	5.316	1.000	-14.426	16.440
GP	1	2	-5.694	9.087	1.000	-32.075	20.686
		3	-4.028	7.870	1.000	-26.874	18.818
		4	-3.819	7.870	1.000	-26.666	19.027
		5	-5.972	7.870	1.000	-28.818	16.874

continued . . .

Table A.41: . . . continued

Inc.	TS att.		MD	SE	p	LB	UB
GP	2	1	5.694	9.087	1.000	-20.686	32.075
		3	1.667	7.870	1.000	-21.179	24.513
		4	1.875	7.870	1.000	-20.971	24.721
		5	-.278	7.870	1.000	-23.124	22.568
	3	1	4.028	7.870	1.000	-18.818	26.874
		2	-1.667	7.870	1.000	-24.513	21.179
		4	.208	6.426	1.000	-18.445	18.862
		5	-1.944	6.426	1.000	-20.598	16.709
	4	1	3.819	7.870	1.000	-19.027	26.666
		2	-1.875	7.870	1.000	-24.721	20.971
		3	-.208	6.426	1.000	-18.862	18.445
		5	-2.153	6.426	1.000	-20.807	16.501
	5	1	5.972	7.870	1.000	-16.874	28.818
		2	.278	7.870	1.000	-22.568	23.124
		3	1.944	6.426	1.000	-16.709	20.598
		4	2.153	6.426	1.000	-16.501	20.807
TS	1	2	-8.542	9.389	1.000	-35.798	18.714
		3	.972	8.131	1.000	-22.632	24.577
		4	-4.167	8.131	1.000	-27.771	19.438
		5	-12.778	8.131	1.000	-36.382	10.827
	2	1	8.542	9.389	1.000	-18.714	35.798
		3	9.514	8.131	1.000	-14.091	33.118
		4	4.375	8.131	1.000	-19.229	27.979
		5	-4.236	8.131	1.000	-27.841	19.368
	3	1	-.972	8.131	1.000	-24.577	22.632
		2	-9.514	8.131	1.000	-33.118	14.091
		4	-5.139	6.639	1.000	-24.412	14.134
		5	-13.750	6.639	.422	-33.023	5.523
	4	1	4.167	8.131	1.000	-19.438	27.771

continued . . .

Table A.41: . . . continued

Inc.	TS att.		MD	SE	p	LB	UB
TS	4	2	-4.375	8.131	1.000	-27.979	19.229
		3	5.139	6.639	1.000	-14.134	24.412
		5	-8.611	6.639	1.000	-27.884	10.662
	5	1	12.778	8.131	1.000	-10.827	36.382
		2	4.236	8.131	1.000	-19.368	27.841
		3	13.750	6.639	.422	-5.523	33.023
		4	8.611	6.639	1.000	-10.662	27.884

Table A.42: Pairwise comparisons based on estimated marginal means for SUS-Total scores with TS attitude as a grouping factor showing patterns of interaction of TS attitude and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

TS att.	Inc.		MD	SE	p	LB	UB
1	SS	GP	4.097	7.676	1.000	-14.751	22.946
		TS	31.458	8.538	.001	10.493	52.424
	GP	SS	-4.097	7.676	1.000	-22.946	14.751
		TS	27.361	7.750	.002	8.331	46.392
	TS	SS	-31.458	8.538	.001	-52.424	-10.493
		GP	-27.361	7.750	.002	-46.392	-8.331
2	SS	GP	2.153	7.676	1.000	-16.696	21.001
		TS	26.667	8.538	.008	5.701	47.632
	GP	SS	-2.153	7.676	1.000	-21.001	16.696
		TS	24.514	7.750	.007	5.483	43.544
	TS	SS	-26.667	8.538	.008	-47.632	-5.701
		GP	-24.514	7.750	.007	-43.544	-5.483
3	SS	GP	-.174	5.428	1.000	-13.502	13.154
		TS	32.187	6.037	$< .001$	17.363	47.012
	GP	SS	.174	5.428	1.000	-13.154	13.502
		TS	32.361	5.480	$< .001$	18.905	45.818

continued . . .

Table A.42: . . . continued

TS att.	Inc.		MD	SE	p	LB	UB
3	TS	SS	-32.187	6.037	$< .001$	-47.012	-17.363
		GP	-32.361	5.480	$< .001$	-45.818	-18.905
4	SS	GP	6.458	5.428	.715	-6.870	19.786
		TS	33.472	6.037	$< .001$	18.647	48.297
	GP	SS	-6.458	5.428	.715	-19.786	6.870
		TS	27.014	5.480	$< .001$	13.557	40.470
	TS	SS	-33.472	6.037	$< .001$	-48.297	-18.647
		GP	-27.014	5.480	$< .001$	-40.470	-13.557
5	SS	GP	5.312	5.428	.994	-8.016	18.641
		TS	25.868	6.037	$< .001$	11.043	40.693
	GP	SS	-5.312	5.428	.994	-18.641	8.016
		TS	20.556	5.480	.001	7.099	34.012
	TS	SS	-25.868	6.037	$< .001$	-40.693	-11.043
		GP	-20.556	5.480	.001	-34.012	-7.099

Table A.43: Post-hoc pairwise comparisons based on observed means with different methods for SUS-Total scores with different TS attitudes. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 149.010$.

Method	TS att.		MD		p		UB
Tukey HSD	1	2	-5.995	5.754	.835	-22.128	10.137
		3	-.938	4.983	1.000	-14.908	13.033
		4	-4.722	4.983	.877	-18.693	9.249
		5	-8.646	4.983	.420	-22.617	5.325
	2	1	5.995	5.754	.835	-10.137	22.128
		3	5.058	4.983	.848	-8.913	19.029
		4	1.273	4.983	.999	-12.698	15.244
		5	-2.650	4.983	.984	-16.621	11.320

continued . . .

Table A.43: . . . continued

Method	TS att.		MD		p		UB
Tukey HSD	3	1	.938	4.983	1.000	-13.033	14.908
		2	-5.058	4.983	.848	-19.029	8.913
		4	-3.785	4.069	.884	-15.192	7.622
		5	-7.708	4.069	.330	-19.116	3.699
	4	1	4.722	4.983	.877	-9.249	18.693
		2	-1.273	4.983	.999	-15.244	12.698
		3	3.785	4.069	.884	-7.622	15.192
		5	-3.924	4.069	.870	-15.331	7.484
	5	1	8.646	4.983	.420	-5.325	22.617
		2	2.650	4.983	.984	-11.320	16.621
		3	7.708	4.069	.330	-3.699	19.116
		4	3.924	4.069	.870	-7.484	15.331
Bonferroni	1	2	-5.995	5.754	1.000	-22.700	10.710
		3	-.938	4.983	1.000	-15.404	13.529
		4	-4.722	4.983	1.000	-19.189	9.745
		5	-8.646	4.983	.874	-23.113	5.821
	2	1	5.995	5.754	1.000	-10.710	22.700
		3	5.058	4.983	1.000	-9.409	19.525
		4	1.273	4.983	1.000	-13.194	15.740
		5	-2.650	4.983	1.000	-17.117	11.816
	3	1	.938	4.983	1.000	-13.529	15.404
		2	-5.058	4.983	1.000	-19.525	9.409
		4	-3.785	4.069	1.000	-15.597	8.027
		5	-7.708	4.069	.625	-19.520	4.104
	4	1	4.722	4.983	1.000	-9.745	19.189
		2	-1.273	4.983	1.000	-15.740	13.194
		3	3.785	4.069	1.000	-8.027	15.597
		5	-3.924	4.069	1.000	-15.736	7.889
	5	1	8.646	4.983	.874	-5.821	23.113

continued . . .

Table A.43: . . . continued

Method	TS att.		MD		p		UB
Bonferroni	5	2	2.650	4.983	1.000	-11.816	17.117
		3	7.708	4.069	.625	-4.104	19.520
		4	3.924	4.069	1.000	-7.889	15.736

SUS-Total with FE

Table A.44: Estimates for SUS-Total scores with FE group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
FE gr.	A	54.57	2.775	49.04	60.11
	B	63.32	2.579	58.18	68.47
	C	60.77	2.138	56.51	65.04
Inc.	SS	72.22	1.801	68.63	75.81
	GP	66.75	2.203	62.36	71.15
	TS	39.70	2.338	35.04	44.36
A *	SS	73.19	3.446	66.32	80.06
	GP	59.87	4.216	51.46	68.28
	TS	30.66	4.475	21.73	39.58
B *	SS	78.15	3.202	71.77	84.54
	GP	69.23	3.918	61.42	77.05
	TS	42.59	4.159	34.29	50.88
C *	SS	65.31	2.655	60.02	70.61
	GP	71.15	3.249	64.67	77.63
	TS	45.86	3.448	38.98	52.74

Table A.45: Pairwise comparisons of main effects of inceptor and FE group, based on estimated marginal means for SUS-Total scores. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
FE gr.	A	B	-8.751	3.788	.071	-18.042	.539
		C	-6.202	3.503	.243	-14.794	2.390
	B	A	8.751	3.788	.071	-.539	18.042
		C	2.549	3.350	1.000	-5.667	10.765
	C	A	6.202	3.503	.243	-2.390	14.794
		B	-2.549	3.350	1.000	-10.765	5.667
Inc.	SS	GP	5.468	2.513	.099	-.698	11.633
		TS	32.518	2.779	$< .001$	25.702	39.334
	GP	SS	-5.468	2.513	.099	-11.633	.698
		TS	27.050	2.781	$< .001$	20.229	33.872
	TS	SS	-32.518	2.779	$< .001$	-39.334	-25.702
		GP	-27.050	2.781	$< .001$	-33.872	-20.229

Table A.46: Pairwise comparisons based on estimated marginal means for SUS-Total scores with FE group as a grouping factor showing patterns of interaction of inceptor and FE group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	FE gr.		MD	SE	p	LB	UB
SS	A	B	-4.963	4.704	.885	-16.502	6.576
		C	7.878	4.350	.223	-2.792	18.549
	B	A	4.963	4.704	.885	-6.576	16.502
		C	12.841	4.160	.009	2.637	23.045
	C	A	-7.878	4.350	.223	-18.549	2.792
		B	-12.841	4.160	.009	-23.045	-2.637
GP	A	B	-9.365	5.756	.325	-23.483	4.754
		C	-11.284	5.323	.113	-24.340	1.772

continued . . .

Table A.46: . . . continued

Inc.	FE gr.		MD	SE	p	LB	UB
GP	B	A	9.365	5.756	.325	-4.754	23.483
		C	-1.919	5.090	1.000	-14.404	10.565
	C	A	11.284	5.323	.113	-1.772	24.340
		B	1.919	5.090	1.000	-10.565	14.404
TS	A	B	-11.927	6.109	.165	-26.912	3.058
		C	-15.201	5.650	.027	-29.059	-1.344
	B	A	11.927	6.109	.165	-3.058	26.912
		C	-3.274	5.402	1.000	-16.526	9.977
	C	A	15.201	5.650	.027	1.344	29.059
		B	3.274	5.402	1.000	-9.977	16.526

Table A.47: Pairwise comparisons based on estimated marginal means for SUS-Total scores with FE group as a grouping factor showing patterns of interaction of FE group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

FE gr.	Inc.		MD	SE	p	LB	UB
A	SS	GP	13.322	4.810	.022	1.524	25.121
		TS	42.533	5.318	$< .001$	29.488	55.578
	GP	SS	-13.322	4.810	.022	-25.121	-1.524
		TS	29.211	5.322	$< .001$	16.156	42.265
	TS	SS	-42.533	5.318	$< .001$	-55.578	-29.488
		GP	-29.211	5.322	$< .001$	-42.265	-16.156
B	SS	GP	8.920	4.470	.150	-2.044	19.885
		TS	35.568	4.942	$< .001$	23.445	47.691
	GP	SS	-8.920	4.470	.150	-19.885	2.044
		TS	26.648	4.946	$< .001$	14.516	38.780
	TS	SS	-35.568	4.942	$< .001$	-47.691	-23.445
		GP	-26.648	4.946	$< .001$	-38.780	-14.516
C	SS	GP	-5.840	3.706	.359	-14.931	3.252

continued . . .

Table A.47: . . . continued

FE gr.	Inc.		MD	SE	*p*	LB	UB
C	SS	TS	19.453	4.098	< .001	9.401	29.505
	GP	SS	5.840	3.706	.359	-3.252	14.931
		TS	25.293	4.101	< .001	15.234	35.352
	TS	SS	-19.453	4.098	< .001	-29.505	-9.401
		GP	-25.293	4.101	< .001	-35.352	-15.234

Table A.48: Post-hoc pairwise comparisons based on observed means with different methods for SUS-Total scores with FE group. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 146.273$.

Method	FE gr.		MD		*p*		UB
Tukey HSD	A	B	-8.751	3.788	.061	-17.822	.319
		C	-6.202	3.503	.187	-14.590	2.185
	B	A	8.751	3.788	.061	-.319	17.822
		C	2.549	3.350	.728	-5.472	10.570
	C	A	6.202	3.503	.187	-2.185	14.590
		B	-2.549	3.350	.728	-10.570	5.472
Bonferroni	A	B	-8.751	3.788	.071	-18.042	.539
		C	-6.202	3.503	.243	-14.794	2.390
	B	A	8.751	3.788	.071	-.539	18.042
		C	2.549	3.350	1.000	-5.667	10.765
	C	A	6.202	3.503	.243	-2.390	14.794
		B	-2.549	3.350	1.000	-10.765	5.667

SUS-Total with VG

Table A.49: Estimates for SUS-Total scores with VG group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
VG gr.	A	61.33	2.051	57.24	65.42
	B	61.73	2.477	56.79	66.67
	C	53.35	3.243	46.88	59.82
Inc.	SS	70.70	2.006	66.70	74.71
	GP	65.28	2.190	60.91	69.65
	TS	40.42	2.569	35.29	45.54
A *	SS	72.23	2.703	66.84	77.62
	GP	69.70	2.950	63.81	75.58
	TS	42.05	3.461	35.15	48.96
B *	SS	71.35	3.264	64.84	77.86
	GP	73.20	3.563	66.10	80.31
	TS	40.63	4.180	32.29	48.96
C *	SS	68.53	4.274	60.00	77.05
	GP	52.95	4.665	43.64	62.25
	TS	38.57	5.473	27.66	49.49

Table A.50: Pairwise comparisons of main effects of inceptor and VG group, based on estimated marginal means for SUS-Total scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
FE gr.	A	B	-.400	3.216	1.000	-8.289	7.489
		C	7.979	3.837	.124	-1.434	17.392
	B	A	.400	3.216	1.000	-7.489	8.289
		C	8.379	4.081	.131	-1.631	18.389
	C	A	-7.979	3.837	.124	-17.392	1.434

continued ...

Table A.50: ... continued

Factor			MD	SE	p	LB	UB
FE gr.	C	B	-8.379	4.081	.131	-18.389	1.631
Inc.	SS	GP	5.422	2.736	.154	-1.288	12.133
		TS	30.288	3.175	$< .001$	22.500	38.076
	GP	SS	-5.422	2.736	.154	-12.133	1.288
		TS	24.865	2.800	$< .001$	17.997	31.734
	TS	SS	-30.288	3.175	$< .001$	-38.076	-22.500
		GP	-24.865	2.800	$< .001$	-31.734	-17.997

Table A.51: Pairwise comparisons based on estimated marginal means for SUS-Total scores with VG group as a grouping factor showing patterns of interaction of inceptor and VG group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	FE gr.		MD	SE	p	LB	UB
SS	A	B	.878	4.238	1.000	-9.518	11.274
		C	3.705	5.057	1.000	-8.700	16.110
	B	A	-.878	4.238	1.000	-11.274	9.518
		C	2.827	5.378	1.000	-10.365	16.020
	C	A	-3.705	5.057	1.000	-16.110	8.700
		B	-2.827	5.378	1.000	-16.020	10.365
GP	A	B	-3.507	4.626	1.000	-14.854	7.840
		C	16.750	5.520	.010	3.211	30.289
	B	A	3.507	4.626	1.000	-7.840	14.854
		C	20.257	5.870	.003	5.858	34.655
	C	A	-16.750	5.520	.010	-30.289	-3.211
		B	-20.257	5.870	.003	-34.655	-5.858
TS	A	B	1.429	5.427	1.000	-11.883	14.741
		C	3.482	6.476	1.000	-12.402	19.366
	B	A	-1.429	5.427	1.000	-14.741	11.883
		C	2.054	6.887	1.000	-14.838	18.946

continued ...

Table A.51: . . . continued

Inc.	FE gr.		MD	SE	p	LB	UB
TS	C	A	-3.482	6.476	1.000	-19.366	12.402
		B	-2.054	6.887	1.000	-18.946	14.838

Table A.52: Pairwise comparisons based on estimated marginal means for SUS-Total scores with VG group as a grouping factor showing patterns of interaction of VG group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

VG gr.	Inc.		MD	SE	p	LB	UB
A	SS	GP	2.536	3.686	1.000	-6.505	11.576
		TS	30.179	4.278	< .001	19.686	40.671
	GP	SS	-2.536	3.686	1.000	-11.576	6.505
		TS	27.643	3.773	< .001	18.389	36.897
	TS	SS	-30.179	4.278	< .001	-40.671	-19.686
		GP	-27.643	3.773	< .001	-36.897	-18.389
B	SS	GP	-1.849	4.451	1.000	-12.766	9.068
		TS	30.729	5.166	< .001	18.059	43.400
	GP	SS	1.849	4.451	1.000	-9.068	12.766
		TS	32.578	4.556	< .001	21.403	43.753
	TS	SS	-30.729	5.166	< .001	-43.400	-18.059
		GP	-32.578	4.556	< .001	-43.753	-21.403
C	SS	GP	15.580	5.828	.028	1.286	29.875
		TS	29.955	6.763	< .001	13.366	46.545
	GP	SS	-15.580	5.828	.028	-29.875	-1.286
		TS	14.375	5.965	.056	-.256	29.006
	TS	SS	-29.955	6.763	< .001	-46.545	-13.366
		GP	-14.375	5.965	.056	-29.006	.256

Table A.53: Post-hoc pairwise comparisons based on observed means with different methods for SUS-Total scores with VG group. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 147.262$.

Method	VG gr.		MD		p		UB
Tukey HSD	A	B	-.400	3.216	.992	-8.101	7.301
		C	7.979	3.837	.102	-1.210	17.168
	B	A	.400	3.216	.992	-7.301	8.101
		C	8.379	4.081	.107	-1.393	18.151
	C	A	-7.979	3.837	.102	-17.168	1.210
		B	-8.379	4.081	.107	-18.151	1.393
Bonferroni	A	B	-.400	3.216	1.000	-8.289	7.489
		C	7.979	3.837	.124	-1.434	17.392
	B	A	.400	3.216	1.000	-7.489	8.289
		C	8.379	4.081	.131	-1.631	18.389
	C	A	-7.979	3.837	.124	-17.392	1.434
		B	-8.379	4.081	.131	-18.389	1.631

SART-D with GP

Table A.54: Estimates for SART-D scores with GP usage as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
GP usage	no	11.54	.436	10.67	12.41
	hardly ever	10.92	.606	9.71	12.13
	sometimes	10.99	.507	9.98	12.00
	a lot	13.22	.654	11.91	14.52
Inc.	SS	10.99	.391	10.21	11.77
	GP	10.66	.397	9.87	11.45

continued . . .

Table A.54: . . . continued

Effect		M	SE	LB	UB
Inc.	TS	13.35	.395	12.56	14.14
no *	SS	10.29	.611	9.07	11.51
	GP	10.68	.622	9.43	11.92
	TS	13.67	.618	12.43	14.90
hardly ever *	SS	9.79	.849	8.09	11.48
	GP	10.02	.864	8.29	11.74
	TS	12.95	.858	11.23	14.66
sometimes *	SS	10.96	.710	9.55	12.38
	GP	10.05	.723	8.61	11.49
	TS	11.95	.718	10.52	13.38
a lot *	SS	12.92	.917	11.09	14.75
	GP	11.90	.933	10.03	13.76
	TS	14.83	.927	12.98	16.68

Table A.55: Pairwise comparisons of main effects of inceptor and GP usage, based on estimated marginal means for SART-D scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
GP usage	no	hardly ever	.627	.747	1.000	-1.401	2.655
		sometimes	.556	.669	1.000	-1.261	2.372
		a lot	-1.672	.786	.222	-3.808	.464
	hardly ever	no	-.627	.747	1.000	-2.655	1.401
		sometimes	-.071	.790	1.000	-2.217	2.075
		a lot	-2.299	.892	.073	-4.721	.124
	sometimes	no	-.556	.669	1.000	-2.372	1.261
		hardly ever	.071	.790	1.000	-2.075	2.217
		a lot	-2.228	.828	.054	-4.476	.021

continued . . .

Table A.55: . . . continued

Factor			MD	SE	p	LB	UB
GP usage	a lot	no	1.672	.786	.222	-.464	3.808
		hardly ever	2.299	.892	.073	-.124	4.721
		sometimes	2.228	.828	.054	-.021	4.476
Inc.	SS	GP	.328	.517	1.000	-.940	1.596
		TS	-2.361	.489	< .001	-3.560	-1.162
	GP	SS	-.328	.517	1.000	-1.596	.940
		TS	-2.689	.441	< .001	-3.771	-1.607
	TS	SS	2.361	.489	< .001	1.162	3.560
		GP	2.689	.441	< .001	1.607	3.771

Table A.56: Pairwise comparisons based on estimated marginal means for SART-D scores with GP usage as a grouping factor showing patterns of interaction of inceptor and GP usage. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	GP usage		MD	SE	p	LB	UB
SS	no	hardly ever	.501	1.046	1.000	-2.341	3.344
		sometimes	-.675	.937	1.000	-3.221	1.871
		a lot	-2.630	1.102	.119	-5.624	.364
	hardly ever	no	-.501	1.046	1.000	-3.344	2.341
		sometimes	-1.177	1.107	1.000	-4.184	1.830
		a lot	-3.131	1.250	.088	-6.526	.264
	sometimes	no	.675	.937	1.000	-1.871	3.221
		hardly ever	1.177	1.107	1.000	-1.830	4.184
		a lot	-1.954	1.160	.580	-5.105	1.197
	a lot	no	2.630	1.102	.119	-.364	5.624
		hardly ever	3.131	1.250	.088	-.264	6.526
		sometimes	1.954	1.160	.580	-1.197	5.105
GP	no	hardly ever	.658	1.065	1.000	-2.234	3.550

continued . . .

Table A.56: ... continued

Inc.	GP usage		MD	SE	p	LB	UB
GP	no	sometimes	.626	.954	1.000	-1.965	3.217
		a lot	-1.220	1.122	1.000	-4.267	1.827
	hardly ever	no	-.658	1.065	1.000	-3.550	2.234
		sometimes	-.032	1.127	1.000	-3.092	3.028
		a lot	-1.878	1.272	.866	-5.333	1.577
	sometimes	no	-.626	.954	1.000	-3.217	1.965
		hardly ever	.032	1.127	1.000	-3.028	3.092
		a lot	-1.846	1.180	.735	-5.053	1.361
	a lot	no	1.220	1.122	1.000	-1.827	4.267
		hardly ever	1.878	1.272	.866	-1.577	5.333
		sometimes	1.846	1.180	.735	-1.361	5.053
TS	no	hardly ever	.720	1.058	1.000	-2.153	3.593
		sometimes	1.717	.947	.446	-.857	4.290
		a lot	-1.167	1.114	1.000	-4.193	1.860
	hardly ever	no	-.720	1.058	1.000	-3.593	2.153
		sometimes	.996	1.119	1.000	-2.043	4.036
		a lot	-1.887	1.263	.839	-5.319	1.545
	sometimes	no	-1.717	.947	.446	-4.290	.857
		hardly ever	-.996	1.119	1.000	-4.036	2.043
		a lot	-2.883	1.173	.099	-6.069	.302
	a lot	no	1.167	1.114	1.000	-1.860	4.193
		hardly ever	1.887	1.263	.839	-1.545	5.319
		sometimes	2.883	1.173	.099	-.302	6.069

Table A.57: Pairwise comparisons based on estimated marginal means for SART-D scores with GP usage as a grouping factor showing patterns of interaction of GP usage and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

GP usage	Inc.		MD	SE	p	LB	UB
no	SS	GP	-.389	.809	1.000	-2.374	1.596
		TS	-3.380	.765	$< .001$	-5.256	-1.503
	GP	SS	.389	.809	1.000	-1.596	2.374
		TS	-2.991	.690	$< .001$	-4.684	-1.297
	TS	SS	3.380	.765	$< .001$	1.503	5.256
		GP	2.991	.690	$< .001$	1.297	4.684
hardly ever	SS	GP	-.232	1.123	1.000	-2.988	2.524
		TS	-3.161	1.062	.012	-5.767	-.555
	GP	SS	.232	1.123	1.000	-2.524	2.988
		TS	-2.929	.959	.010	-5.281	-.576
	TS	SS	3.161	1.062	.012	.555	5.767
		GP	2.929	.959	.010	.576	5.281
sometimes	SS	GP	.913	.940	1.000	-1.393	3.218
		TS	-.987	.889	.811	-3.168	1.193
	GP	SS	-.913	.940	1.000	-3.218	1.393
		TS	-1.900	.802	.062	-3.868	.068
	TS	SS	.987	.889	.811	-1.193	3.168
		GP	1.900	.802	.062	-.068	3.868
a lot	SS	GP	1.021	1.213	1.000	-1.956	3.998
		TS	-1.917	1.147	.298	-4.732	.898
	GP	SS	-1.021	1.213	1.000	-3.998	1.956
		TS	-2.938	1.035	.018	-5.478	-.397
	TS	SS	1.917	1.147	.298	-.898	4.732
		GP	2.938	1.035	.018	.397	5.478

Table A.58: Post-hoc pairwise comparisons based on observed means with different methods for SART-D scores with GP usage. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 5.139$.

Method	GP usage		MD	SE	p	LB	UB
Tukey HSD	no	hardly ever	.627	.747	.836	-1.339	2.592
		sometimes	.556	.669	.840	-1.205	2.316
		a lot	-1.672	.786	.155	-3.743	.399
	hardly ever	no	-.627	.747	.836	-2.592	1.339
		sometimes	-.071	.790	1.000	-2.150	2.009
		a lot	-2.299	.892	.057	-4.646	.049
	sometimes	no	-.556	.669	.840	-2.316	1.205
		hardly ever	.071	.790	1.000	-2.009	2.150
		a lot	-2.228	.828	.043	-4.407	-.049
	a lot	no	1.672	.786	.155	-.399	3.743
		hardly ever	2.299	.892	.057	-.049	4.646
		sometimes	2.228	.828	.043	.049	4.407
Bonferroni	no	hardly ever	.627	.747	1.000	-1.401	2.655
		sometimes	.556	.669	1.000	-1.261	2.372
		a lot	-1.672	.786	.222	-3.808	.464
	hardly ever	no	-.627	.747	1.000	-2.655	1.401
		sometimes	-.071	.790	1.000	-2.217	2.075
		a lot	-2.299	.892	.073	-4.721	.124
	sometimes	no	-.556	.669	1.000	-2.372	1.261
		hardly ever	.071	.790	1.000	-2.075	2.217
		a lot	-2.228	.828	.054	-4.476	.021
	a lot	no	1.672	.786	.222	-.464	3.808
		hardly ever	2.299	.892	.073	-.124	4.721
		sometimes	2.228	.828	.054	-.021	4.476

SART-D with MG

Table A.59: Estimates for SART-D scores with MG usage as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
MG	no / h. ever	11.71	.433	10.85	12.57
	used to	12.13	.521	11.09	13.17
	yes	10.86	.476	9.91	11.81
Inc.	SS	10.92	.383	10.15	11.68
	GP	10.52	.369	9.79	11.26
	TS	13.26	.390	12.48	14.04
no / h. ever *	SS	10.04	.601	8.84	11.24
	GP	11.55	.579	10.40	12.71
	TS	13.53	.612	12.31	14.75
used to *	SS	11.85	.724	10.41	13.29
	GP	10.80	.697	9.41	12.19
	TS	13.75	.737	12.28	15.22
yes *	SS	10.86	.661	9.55	12.18
	GP	9.22	.636	7.95	10.49
	TS	12.49	.673	11.15	13.83

Table A.60: Pairwise comparisons of main effects of inceptor and MG usage, based on estimated marginal means for SART-D scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
MG	no / h. ever	used to	-.424	.678	1.000	-2.086	1.239
		yes	.852	.644	.569	-.726	2.431
	used to	no / h. ever	.424	.678	1.000	-1.239	2.086
		yes	1.276	.706	.225	-.456	3.008
	yes	no / h. ever	-.852	.644	.569	-2.431	.726

continued ...

Table A.60: . . . continued

Factor			MD	SE	p	LB	UB
MG	yes	used to	-1.276	.706	.225	-3.008	.456
Inc.	SS	GP	.396	.469	1.000	-.755	1.547
		TS	-2.339	.472	$< .001$	-3.496	-1.181
	GP	SS	-.396	.469	1.000	-1.547	.755
		TS	-2.735	.420	$< .001$	-3.764	-1.705
	TS	SS	2.339	.472	$< .001$	1.181	3.496
		GP	2.735	.420	$< .001$	1.705	3.764

Table A.61: Pairwise comparisons based on estimated marginal means for SART-D scores with MG usage as a grouping factor showing patterns of interaction of inceptor and MG usage. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	MG		MD	SE	p	LB	UB
SS	no / h. ever	used to	-1.807	.941	.177	-4.115	.501
		yes	-.821	.893	1.000	-3.012	1.370
	used to	no / h. ever	1.807	.941	.177	-.501	4.115
		yes	.985	.980	.954	-1.418	3.389
	yes	no / h. ever	.821	.893	1.000	-1.370	3.012
		used to	-.985	.980	.954	-3.389	1.418
GP	no / h. ever	used to	.752	.906	1.000	-1.469	2.973
		yes	2.333	.860	.025	.224	4.442
	used to	no / h. ever	-.752	.906	1.000	-2.973	1.469
		yes	1.581	.943	.294	-.732	3.895
	yes	no / h. ever	-2.333	.860	.025	-4.442	-.224
		used to	-1.581	.943	.294	-3.895	.732
TS	no / h. ever	used to	-.216	.958	1.000	-2.565	2.134
		yes	1.045	.909	.763	-1.186	3.275
	used to	no / h. ever	.216	.958	1.000	-2.134	2.565
		yes	1.260	.998	.632	-1.187	3.708

continued . . .

Table A.61: . . . continued

Inc.	MG		MD	SE	p	LB	UB
TS	yes	no / h. ever	-1.045	.909	.763	-3.275	1.186
		used to	-1.260	.998	.632	-3.708	1.187

Table A.62: Pairwise comparisons based on estimated marginal means for SART-D scores with MG usage as a grouping factor showing patterns of interaction of MG usage and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

MG	Inc.		MD	SE	p	LB	UB
no / h. ever	SS	GP	-1.509	.736	.132	-3.314	.297
		TS	-3.491	.740	< .001	-5.307	-1.676
	GP	SS	1.509	.736	.132	-.297	3.314
		TS	-1.983	.658	.011	-3.598	-.368
	TS	SS	3.491	.740	< .001	1.676	5.307
		GP	1.983	.658	.011	.368	3.598
used to	SS	GP	1.050	.886	.720	-1.124	3.224
		TS	-1.900	.891	.110	-4.087	.287
	GP	SS	-1.050	.886	.720	-3.224	1.124
		TS	-2.950	.793	.001	-4.895	-1.005
	TS	SS	1.900	.891	.110	-.287	4.087
		GP	2.950	.793	.001	1.005	4.895
yes	SS	GP	1.646	.809	.137	-.339	3.630
		TS	-1.625	.814	.149	-3.621	.371
	GP	SS	-1.646	.809	.137	-3.630	.339
		TS	-3.271	.724	< .001	-5.046	-1.495
	TS	SS	1.625	.814	.149	-.371	3.621
		GP	3.271	.724	< .001	1.495	5.046
a lot	SS	GP	1.021	1.213	1.000	-1.956	3.998
		TS	-1.917	1.147	.298	-4.732	.898
	GP	SS	-1.021	1.213	1.000	-3.998	1.956

continued . . .

Table A.62: . . . continued

MG	Inc.		MD	SE	p	LB	UB
		TS	-2.938	1.035	.018	-5.478	-.397
	TS	SS	1.917	1.147	.298	-.898	4.732
		GP	2.938	1.035	.018	.397	5.478

Table A.63: Post-hoc pairwise comparisons based on observed means with different methods for SART-D scores with MG usage. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 5.439$.

Method	MG		MD		p		UB
Tukey HSD	no / h. ever	used to	-.424	.678	.807	-2.047	1.200
		yes	.852	.644	.387	-.689	2.393
	used to	no / h. ever	.424	.678	.807	-1.200	2.047
		yes	1.276	.706	.175	-.415	2.966
	yes	no / h. ever	-.852	.644	.387	-2.393	.689
		used to	-1.276	.706	.175	-2.966	.415
Bonferroni	no / h. ever	used to	-.424	.678	1.000	-2.086	1.239
		yes	.852	.644	.569	-.726	2.431
	used to	no / h. ever	.424	.678	1.000	-1.239	2.086
		yes	1.276	.706	.225	-.456	3.008
	yes	no / h. ever	-.852	.644	.569	-2.431	.726
		used to	-1.276	.706	.225	-3.008	.456

NASA-TLX with VG

Table A.64: Estimates for NASA-TLX scores with VG group. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the VG group and the inceptor.

Effect		M	SE	LB	UB
VG gr.	A	49.95	1.945	46.07	53.83
	B	47.34	2.315	42.72	51.96
	C	56.50	3.031	50.45	62.54
Inceptor	SS	49.49	1.843	45.81	53.16
	GP	43.88	1.903	40.09	47.68
	TS	60.41	2.009	56.40	64.42
A *	SS	48.02	2.511	43.01	53.03
	GP	42.18	2.594	37.00	47.35
	TS	59.64	2.738	54.18	65.10
B *	SS	44.24	2.989	38.28	50.21
	GP	37.44	3.088	31.28	43.60
	TS	60.33	3.259	53.83	66.83
C *	SS	56.20	3.914	48.39	64.01
	GP	52.03	4.043	43.97	60.10
	TS	61.26	4.267	52.75	69.77

Table A.65: Pairwise comparisons of main effects of inceptor and VG group based on estimated marginal means for NASA-TLX scores. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
VG gr.	A	B	2.608	3.023	1.000	-4.811	10.026
		C	-6.552	3.601	.220	-15.389	2.284
	B	A	-2.608	3.023	1.000	-10.026	4.811
		C	-9.160	3.814	.057	-18.518	.198
	C	A	6.552	3.601	.220	-2.284	15.389

continued ...

Table A.65: . . . continued

Factor			MD	SE	p	LB	UB
VG gr.	C	B	9.160	3.814	.057	-.198	18.518
Inc	SS	GP	5.603	2.132	.032	.371	10.835
		TS	-10.923	2.217	$< .001$	-16.362	-5.484
	GP	SS	-5.603	2.132	.032	-10.835	-.371
		TS	-16.526	2.319	$< .001$	-22.216	-10.835
	TS	SS	10.923	2.217	$< .001$	5.484	16.362
		GP	16.526	2.319	$< .001$	10.835	22.216

Table A.66: Pairwise comparisons based on estimated marginal means for NASA-TLX scores showing patterns of interaction of Inceptor and VG group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	VG gr.		MD	SE	p	LB	UB
SS	A	B	3.777	3.904	1.000	-5.803	13.357
		C	-8.182	4.650	.249	-19.592	3.229
	B	A	-3.777	3.904	1.000	-13.357	5.803
		C	-11.958	4.925	.053	-24.043	.126
	C	A	8.182	4.650	.249	-3.229	19.592
		B	11.958	4.925	.053	-.126	24.043
GP	A	B	4.735	4.033	.733	-5.160	14.630
		C	-9.856	4.803	.132	-21.642	1.930
	B	A	-4.735	4.033	.733	-14.630	5.160
		C	-14.592	5.087	.016	-27.073	-2.110
	C	A	9.856	4.803	.132	-1.930	21.642
		B	14.592	5.087	.016	2.110	27.073
TS	A	B	-.689	4.257	1.000	-11.134	9.756
		C	-1.619	5.070	1.000	-14.060	10.822
	B	A	.689	4.257	1.000	-9.756	11.134
		C	-.930	5.369	1.000	-14.105	12.245
	C	A	1.619	5.070	1.000	-10.822	14.060

continued . . .

Table A.66: . . . continued

Inc.	VG gr.		MD	SE	p	LB	UB
TS	C	B	.930	5.369	1.000	-12.245	14.105

Table A.67: Pairwise comparisons based on estimated marginal means for NASA-TLX scores showing patterns of interaction of VG group and Inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

VG gr.	Inc.		MD	SE	p	LB	UB
A	SS	GP	5.842	2.906	.145	-1.289	12.972
		TS	-11.622	3.021	.001	-19.035	-4.209
	GP	SS	-5.842	2.906	.145	-12.972	1.289
		TS	-17.463	3.161	< .001	-25.218	-9.708
	TS	SS	11.622	3.021	.001	4.209	19.035
		GP	17.463	3.161	< .001	9.708	25.218
B	SS	GP	6.800	3.459	.160	-1.687	15.287
		TS	-16.088	3.596	< .001	-24.911	-7.265
	GP	SS	-6.800	3.459	.160	-15.287	1.687
		TS	-22.888	3.762	< .001	-32.118	-13.657
	TS	SS	16.088	3.596	< .001	7.265	24.911
		GP	22.888	3.762	< .001	13.657	32.118
C	SS	GP	4.167	4.529	1.000	-6.946	15.279
		TS	-5.060	4.708	.859	-16.612	6.493
	GP	SS	-4.167	4.529	1.000	-15.279	6.946
		TS	-9.226	4.925	.196	-21.312	2.859
	TS	SS	5.060	4.708	.859	-6.493	16.612
		GP	9.226	4.925	.196	-2.859	21.312

Table A.68: Post-hoc pairwise comparisons based on observed means with different methods for NASA-TLX with VG group. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 128.605$.

Method	VG gr.		MD	SE	p	LB	UB
Tukey HSD	A	B	2.608	3.023	.666	-4.634	9.850
		C	-6.552	3.601	.171	-15.178	2.073
	B	A	-2.608	3.023	.666	-9.850	4.634
		C	-9.160	3.814	.049	-18.295	-.025
	C	A	6.552	3.601	.171	-2.073	15.178
		B	9.160	3.814	.049	.025	18.295
Bonferroni	A	B	2.608	3.023	1.000	-4.811	10.026
		C	-6.552	3.601	.220	-15.389	2.284
	B	A	-2.608	3.023	1.000	-10.026	4.811
		C	-9.160	3.814	.057	-18.518	.198
	C	A	6.552	3.601	.220	-2.284	15.389
		B	9.160	3.814	.057	-.198	18.518

ANOVA PS DRV with gender

Table A.69: Estimates for PS in DRV scenario with gender as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Gender	Female	51.50	3.768	43.98	59.01
	Male	66.89	2.175	62.56	71.23
	Pref. not to say	80.11	15.986	48.23	112.00
Inc.	SS	70.49	7.379	55.77	85.21
	GP	78.84	6.779	65.32	92.36
	TS	49.18	5.836	37.54	60.82

continued ...

Table A.69: . . . continued

Effect		M	SE	LB	UB
Female	* SS	52.98	5.035	42.94	63.02
	* GP	63.50	4.625	54.28	72.72
	* TS	38.02	3.982	30.08	45.96
Male	* SS	70.46	2.907	64.66	76.26
	* GP	77.08	2.670	71.76	82.41
	* TS	53.14	2.299	48.56	57.73
Pref. not to say	* SS	88.03	21.361	45.43	130.63
	* GP	95.93	19.622	56.79	135.06
	* TS	56.39	16.894	22.69	90.08

Table A.70: Pairwise comparisons of main effects of inceptor and gender, based on estimated marginal means for PS in DRV scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Gender	Female	Male	-15.396	4.351	.002	-26.068	-4.724
		Pref. not to say	-28.616	16.424	.258	-68.901	11.670
	Male	Female	15.396	4.351	.002	4.724	26.068
		Pref. not to say	-13.219	16.133	1.000	-52.792	26.353
	Pref. not to say	Female	28.616	16.424	.258	-11.670	68.901
		Male	13.219	16.133	1.000	-26.353	52.792
Inc.	SS	GP	-8.345	5.891	.483	-22.795	6.104
		TS	21.308	6.717	.007	4.833	37.783
	GP	SS	8.345	5.891	.483	-6.104	22.795
		TS	29.653	7.008	$< .001$	12.464	46.842
	TS	SS	-21.308	6.717	.007	-37.783	-4.833
		GP	-29.653	7.008	$< .001$	-46.842	-12.464

Table A.71: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with gender as a grouping factor showing patterns of interaction of inceptor and gender. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	Female	Male	-17.481	5.814	.011	-31.741	-3.220
		Pref. not to say	-35.050	21.946	.344	-88.881	18.781
	Male	Female	17.481	5.814	.011	3.220	31.741
		Pref. not to say	-17.569	21.558	1.000	-70.448	35.309
	Pref. not to say	Female	35.050	21.946	.344	-18.781	88.881
		Male	17.569	21.558	1.000	-35.309	70.448
GP	Female	Male	-13.582	5.341	.040	-26.682	-.482
		Pref. not to say	-32.426	20.160	.337	-81.877	17.024
	Male	Female	13.582	5.341	.040	.482	26.682
		Pref. not to say	-18.844	19.803	1.000	-67.419	29.731
	Pref. not to say	Female	32.426	20.160	.337	-17.024	81.877
		Male	18.844	19.803	1.000	-29.731	67.419
TS	Female	Male	-15.126	4.598	.005	-26.404	-3.848
		Pref. not to say	-18.371	17.357	.880	-60.944	24.203
	Male	Female	15.126	4.598	.005	3.848	26.404
		Pref. not to say	-3.244	17.049	1.000	-45.064	38.575
	Pref. not to say	Female	18.371	17.357	.880	-24.203	60.944
		Male	3.244	17.049	1.000	-38.575	45.064

Table A.72: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with gender as a grouping factor showing patterns of interaction of gender and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
Female	SS	GP	-10.519	4.019	.033	-20.378	-.661

continued . . .

Table A.72: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
Female	SS	TS	14.963	4.583	.005	3.723	26.204
	GP	SS	10.519	4.019	.033	.661	20.378
		TS	25.483	4.781	< .001	13.755	37.211
	TS	SS	-14.963	4.583	.005	-26.204	-3.723
		GP	-25.483	4.781	< .001	-37.211	-13.755
Male	SS	GP	-6.621	2.320	.017	-12.313	-.929
		TS	17.318	2.646	< .001	10.828	23.807
	GP	SS	6.621	2.320	.017	.929	12.313
		TS	23.938	2.760	< .001	17.167	30.709
	TS	SS	-17.318	2.646	< .001	-23.807	-10.828
		GP	-23.938	2.760	< .001	-30.709	-17.167
Pref. not to say	SS	GP	-7.896	17.052	1.000	-49.722	33.931
		TS	31.643	19.443	.324	-16.048	79.333
	GP	SS	7.896	17.052	1.000	-33.931	49.722
		TS	39.538	20.285	.166	-10.218	89.295
	TS	SS	-31.643	19.443	.324	-79.333	16.048
		GP	-39.538	20.285	.166	-89.295	10.218

ANOVA PS DRV with hand

Table A.73: Estimates for PS in DRV scenario with handedness as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Handedness	Ambidextrous	81.36	9.844	61.73	100.99
	Left-handed	62.96	6.445	50.10	75.81
	Right-handed	62.45	2.148	58.17	66.74
Inc.	SS	68.68	5.258	58.20	79.17
	GP	80.83	4.736	71.38	90.28

continued . . .

Table A.73: . . . continued

Effect		M	SE	LB	UB
Inc.	TS	57.26	4.106	49.07	65.45
Ambidextrous	* SS	81.30	12.983	55.40	107.19
	* GP	92.69	11.695	69.36	116.01
	* TS	70.10	10.137	49.88	90.32
Left-handed	* SS	58.16	8.499	41.21	75.11
	* GP	77.04	7.656	61.77	92.31
	* TS	53.67	6.636	40.44	66.91
Right-handed	* SS	66.60	2.833	60.95	72.25
	* GP	72.76	2.552	67.67	77.85
	* TS	48.01	2.212	43.59	52.42

Table A.74: Pairwise comparisons of main effects of inceptor and hand, based on estimated marginal means for PS in DRV scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Handedness	Ambidextrous	Left-handed	18.402	11.766	.367	-10.459	47.263
		Right-handed	18.905	10.076	.194	-5.810	43.620
	Left-handed	Ambidextrous	-18.402	11.766	.367	-47.263	10.459
		Right-handed	.503	6.793	1.000	-16.160	17.166
	Right-handed	Ambidextrous	-18.905	10.076	.194	-43.620	5.810
		Left-handed	-.503	6.793	1.000	-17.166	16.160
Inc.	SS	GP	-12.144	3.902	.008	-21.716	-2.573
		TS	11.426	4.454	.037	.500	22.351
	GP	SS	12.144	3.902	.008	2.573	21.716
		TS	23.570	4.763	$< .001$	11.888	35.252
	TS	SS	-11.426	4.454	.037	-22.351	-.500
		GP	-23.570	4.763	$< .001$	-35.252	-11.888

Table A.75: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with handedness as a grouping factor showing patterns of interaction of inceptor and hand. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	Ambidextrous	Left-handed	23.135	15.517	.421	-14.927	61.198
		Right-handed	14.701	13.288	.817	-17.894	47.296
	Left-handed	Ambidextrous	-23.135	15.517	.421	-61.198	14.927
		Right-handed	-8.434	8.959	1.000	-30.409	13.541
	Right-handed	Ambidextrous	-14.701	13.288	.817	-47.296	17.894
		Left-handed	8.434	8.959	1.000	-13.541	30.409
GP	Ambidextrous	Left-handed	15.645	13.978	.801	-18.641	49.932
		Right-handed	19.924	11.970	.301	-9.437	49.285
	Left-handed	Ambidextrous	-15.645	13.978	.801	-49.932	18.641
		Right-handed	4.279	8.070	1.000	-15.517	24.074
	Right-handed	Ambidextrous	-19.924	11.970	.301	-49.285	9.437
		Left-handed	-4.279	8.070	1.000	-24.074	15.517
TS	Ambidextrous	Left-handed	16.426	12.116	.539	-13.294	46.146
		Right-handed	22.091	10.376	.110	-3.359	47.542
	Left-handed	Ambidextrous	-16.426	12.116	.539	-46.146	13.294
		Right-handed	5.665	6.995	1.000	-11.493	22.824
	Right-handed	Ambidextrous	-22.091	10.376	.110	-47.542	3.359
		Left-handed	-5.665	6.995	1.000	-22.824	11.493

Table A.76: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with handedness as a grouping factor showing patterns of interaction of handedness and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
Ambidextrous	SS	GP	-11.389	9.635	.724	-35.023	12.245

continued . . .

Table A.76: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
Ambidextrous	SS	TS	11.199	10.998	.936	-15.778	38.175
	GP	SS	11.389	9.635	.724	-12.245	35.023
		TS	22.588	11.760	.176	-6.257	51.432
	TS	SS	-11.199	10.998	.936	-38.175	15.778
		GP	-22.588	11.760	.176	-51.432	6.257
Left-handed	SS	GP	-18.879	6.308	.011	-34.351	-3.407
		TS	4.490	7.200	1.000	-13.171	22.150
	GP	SS	18.879	6.308	.011	3.407	34.351
		TS	23.368	7.698	.010	4.485	42.252
	TS	SS	-4.490	7.200	1.000	-22.150	13.171
		GP	-23.368	7.698	.010	-42.252	-4.485
Right-handed	SS	GP	-6.166	2.103	.014	-11.323	-1.008
		TS	18.589	2.400	$< .001$	12.702	24.476
	GP	SS	6.166	2.103	.014	1.008	11.323
		TS	24.755	2.566	$< .001$	18.460	31.049
	TS	SS	-18.589	2.400	$< .001$	-24.476	-12.702
		GP	-24.755	2.566	$< .001$	-31.049	-18.460

Table A.77: Post-hoc pairwise comparisons based on observed means with different methods for PS in DRV scenario with handedness grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 290.730$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	Ambidextrous	Left-handed	18.402	11.766	.268	-9.773	46.577
		Right-handed	18.905	10.076	.153	-5.222	43.033
	Left-handed	Ambidextrous	-18.402	11.766	.268	-46.577	9.773
		Right-handed	.503	6.793	.997	-15.763	16.770
	Right-handed	Ambidextrous	-18.905	10.076	.153	-43.033	5.222
		Left-handed	-.503	6.793	.997	-16.770	15.763

continued . . .

Table A.77: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Bonferroni	Ambidextrous	Left-handed	18.402	11.766	.367	-10.459	47.263
		Right-handed	18.905	10.076	.194	-5.810	43.620
	Left-handed	Ambidextrous	-18.402	11.766	.367	-47.263	10.459
		Right-handed	.503	6.793	1.000	-16.160	17.166
	Right-handed	Ambidextrous	-18.905	10.076	.194	-43.620	5.810
		Left-handed	-.503	6.793	1.000	-17.166	16.160

ANOVA PS DRV with VG frequency

Table A.78: Estimates for PS in DRV scenario with VG frequency as a grouping factor. hpw - hours per week; M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
VG freq.	no / h. ever	60.25	4.428	51.41	69.09
	used to, > 3 hpw	57.59	3.532	50.54	64.63
	used to, < 3 hpw	59.44	4.996	49.47	69.41
	yes, > 3 hpw	71.97	4.595	62.80	81.14
	yes, < 3 hpw	70.74	4.595	61.57	79.91
Inc.	SS	67.47	2.597	62.29	72.65
	GP	74.92	2.378	70.18	79.66
	TS	49.60	2.143	45.32	53.87
no / h. ever	SS	64.21	5.770	52.70	75.72
	GP	74.19	5.283	63.64	84.73
	TS	42.36	4.761	32.85	51.86
used to, > 3 hpw	SS	57.75	4.603	48.57	66.94
	GP	66.26	4.215	57.85	74.67
	TS	48.74	3.798	41.16	56.32
used to, < 3 hpw	SS	61.58	6.510	48.59	74.57

continued . . .

Table A.78: . . . continued

Effect		M	SE	LB	UB
used to, < 3 hpw	GP	69.51	5.960	57.62	81.40
	TS	47.22	5.371	36.50	57.94
yes, > 3 hpw	SS	77.23	5.988	65.28	89.18
	GP	83.19	5.483	72.25	94.13
	TS	55.49	4.941	45.63	65.35
yes, < 3 hpw	SS	76.58	5.988	64.63	88.53
	GP	81.45	5.483	70.51	92.39
	TS	54.18	4.941	44.32	64.04

Table A.79: Pairwise comparisons of main effects of inceptor and VG frequency, based on estimated marginal means for PS in DRV scenario. hpw - hours per week; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
VG freq.	no / h. ever	used to, > 3 hpw	2.665	5.664	1.000	-13.770	19.100
		used to, < 3 hpw	.813	6.676	1.000	-18.556	20.182
		yes, > 3 hpw	-11.720	6.381	.706	-30.236	6.796
		yes, < 3 hpw	-10.487	6.381	1.000	-29.003	8.029
	used to, > 3 hpw	no / h. ever	-2.665	5.664	1.000	-19.100	13.770
		used to, < 3 hpw	-1.852	6.118	1.000	-19.604	15.901
		yes, > 3 hpw	-14.385	5.796	.155	-31.202	2.432
		yes, < 3 hpw	-13.151	5.796	.264	-29.969	3.666
	used to, < 3 hpw	no / h. ever	-.813	6.676	1.000	-20.182	18.556
		used to, > 3 hpw	1.852	6.118	1.000	-15.901	19.604
		yes, > 3 hpw	-12.533	6.788	.692	-32.227	7.161
		yes, < 3 hpw	-11.300	6.788	1.000	-30.994	8.394

continued . . .

Table A.79: . . . continued

Factor			MD	SE	*p*	LB	UB
VG freq.	yes, > 3 hpw	no / h. ever	11.720	6.381	.706	-6.796	30.236
		used to, > 3 hpw	14.385	5.796	.155	-2.432	31.202
		used to, < 3 hpw	12.533	6.788	.692	-7.161	32.227
		yes, < 3 hpw	1.233	6.499	1.000	-17.622	20.089
	yes, < 3 hpw	no / h. ever	10.487	6.381	1.000	-8.029	29.003
		used to, > 3 hpw	13.151	5.796	.264	-3.666	29.969
		used to, < 3 hpw	11.300	6.788	1.000	-8.394	30.994
		yes, > 3 hpw	-1.233	6.499	1.000	-20.089	17.622
Inc.	SS	GP	-7.447	2.080	.002	-12.552	-2.343
		TS	17.876	2.272	< .001	12.298	23.453
	GP	SS	7.447	2.080	.002	2.343	12.552
		TS	25.323	2.395	< .001	19.445	31.200
	TS	SS	-17.876	2.272	< .001	-23.453	-12.298
		GP	-25.323	2.395	< .001	-31.200	-19.445

Table A.80: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with VG frequency as a grouping factor showing patterns of interaction of inceptor and VG frequency. Inc. - inceptor; hpw - hours per week; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	*p*	LB	UB
SS	no / h. ever	used to, > 3 hpw	6.457	7.381	1.000	-14.961	27.874
		used to, < 3 hpw	2.626	8.699	1.000	-22.614	27.867
		yes, > 3 hpw	-13.023	8.316	1.000	-37.152	11.106
		yes, < 3 hpw	-12.373	8.316	1.000	-36.502	11.756
	used to, > 3 hpw	no / h. ever	-6.457	7.381	1.000	-27.874	14.961
		used to, < 3 hpw	-3.830	7.973	1.000	-26.964	19.303
		yes, > 3 hpw	-19.480	7.553	.121	-41.395	2.435
		yes, < 3 hpw	-18.830	7.553	.151	-40.745	3.085

continued . . .

Table A.80: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	used to, < 3 hpw	no / h. ever	-2.626	8.699	1.000	-27.867	22.614
		used to, > 3 hpw	3.830	7.973	1.000	-19.303	26.964
		yes, > 3 hpw	-15.649	8.845	.813	-41.314	10.015
		yes, < 3 hpw	-14.999	8.845	.945	-40.664	10.665
	yes, > 3 hpw	no / h. ever	13.023	8.316	1.000	-11.106	37.152
		used to, > 3 hpw	19.480	7.553	.121	-2.435	41.395
		used to, < 3 hpw	15.649	8.845	.813	-10.015	41.314
		yes, < 3 hpw	.650	8.469	1.000	-23.922	25.222
	yes, < 3 hpw	no / h. ever	12.373	8.316	1.000	-11.756	36.502
		used to, > 3 hpw	18.830	7.553	.151	-3.085	40.745
		used to, < 3 hpw	14.999	8.845	.945	-10.665	40.664
		yes, > 3 hpw	-.650	8.469	1.000	-25.222	23.922
GP	no / h. ever	used to, > 3 hpw	7.925	6.758	1.000	-11.685	27.534
		used to, < 3 hpw	4.676	7.965	1.000	-18.434	27.786
		yes, > 3 hpw	-9.000	7.614	1.000	-31.092	13.092
		yes, < 3 hpw	-7.264	7.614	1.000	-29.356	14.828
	used to, > 3 hpw	no / h. ever	-7.925	6.758	1.000	-27.534	11.685
		used to, < 3 hpw	-3.249	7.300	1.000	-24.429	17.932
		yes, > 3 hpw	-16.925	6.915	.170	-36.990	3.140
		yes, < 3 hpw	-15.188	6.915	.315	-35.253	4.877
	used to, < 3 hpw	no / h. ever	-4.676	7.965	1.000	-27.786	18.434
		used to, > 3 hpw	3.249	7.300	1.000	-17.932	24.429
		yes, > 3 hpw	-13.677	8.098	.958	-37.174	9.821
		yes, < 3 hpw	-11.940	8.098	1.000	-35.438	11.558
	yes, > 3 hpw	no / h. ever	9.000	7.614	1.000	-13.092	31.092
		used to, > 3 hpw	16.925	6.915	.170	-3.140	36.990
		used to, < 3 hpw	13.677	8.098	.958	-9.821	37.174
		yes, < 3 hpw	1.737	7.754	1.000	-20.761	24.234

continued . . .

Table A.80: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
GP	yes, < 3 hpw	no / h. ever	7.264	7.614	1.000	-14.828	29.356
		used to, > 3 hpw	15.188	6.915	.315	-4.877	35.253
		used to, < 3 hpw	11.940	8.098	1.000	-11.558	35.438
		yes, > 3 hpw	-1.737	7.754	1.000	-24.234	20.761
TS	no / h. ever	used to, > 3 hpw	-6.387	6.090	1.000	-24.058	11.284
		used to, < 3 hpw	-4.863	7.178	1.000	-25.688	15.963
		yes, > 3 hpw	-13.136	6.861	.598	-33.045	6.772
		yes, < 3 hpw	-11.823	6.861	.894	-31.731	8.085
	used to, > 3 hpw	no / h. ever	6.387	6.090	1.000	-11.284	24.058
		used to, < 3 hpw	1.524	6.578	1.000	-17.563	20.612
		yes, > 3 hpw	-6.749	6.232	1.000	-24.831	11.333
		yes, < 3 hpw	-5.436	6.232	1.000	-23.518	12.646
	used to, < 3 hpw	no / h. ever	4.863	7.178	1.000	-15.963	25.688
		used to, > 3 hpw	-1.524	6.578	1.000	-20.612	17.563
		yes, > 3 hpw	-8.274	7.298	1.000	-29.449	12.902
		yes, < 3 hpw	-6.960	7.298	1.000	-28.136	14.215
	yes, > 3 hpw	no / h. ever	13.136	6.861	.598	-6.772	33.045
		used to, > 3 hpw	6.749	6.232	1.000	-11.333	24.831
		used to, < 3 hpw	8.274	7.298	1.000	-12.902	29.449
		yes, < 3 hpw	1.313	6.987	1.000	-18.961	21.587
	yes, < 3 hpw	no / h. ever	11.823	6.861	.894	-8.085	31.731
		used to, > 3 hpw	5.436	6.232	1.000	-12.646	23.518
		used to, < 3 hpw	6.960	7.298	1.000	-14.215	28.136
		yes, > 3 hpw	-1.313	6.987	1.000	-21.587	18.961

Table A.81: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with VG frequency as a grouping factor showing patterns of interaction of VG frequency and inceptor. Inc. - inceptor; hpw - hours per week; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	*p*	LB	UB
no / h. ever	SS	GP	-9.977	4.621	.103	-21.320	1.366
		TS	21.855	5.049	< .001	9.460	34.249
	GP	SS	9.977	4.621	.103	-1.366	21.320
		TS	31.832	5.321	< .001	18.771	44.893
	TS	SS	-21.855	5.049	< .001	-34.249	-9.460
		GP	-31.832	5.321	< .001	-44.893	-18.771
used to, > 3 hpw	SS	GP	-8.509	3.686	.072	-17.558	.539
		TS	9.011	4.028	.086	-.877	18.898
	GP	SS	8.509	3.686	.072	-.539	17.558
		TS	17.520	4.245	< .001	7.101	27.939
	TS	SS	-9.011	4.028	.086	-18.898	.877
		GP	-17.520	4.245	< .001	-27.939	-7.101
used to, < 3 hpw	SS	GP	-7.927	5.213	.399	-20.724	4.869
		TS	14.366	5.697	.042	.383	28.349
	GP	SS	7.927	5.213	.399	-4.869	20.724
		TS	22.293	6.003	.001	7.558	37.028
	TS	SS	-14.366	5.697	.042	-28.349	-.383
		GP	-22.293	6.003	.001	-37.028	-7.558
yes, > 3 hpw	SS	GP	-5.954	4.795	.656	-17.725	5.817
		TS	21.742	5.240	< .001	8.879	34.604
	GP	SS	5.954	4.795	.656	-5.817	17.725
		TS	27.696	5.522	< .001	14.142	41.250
	TS	SS	-21.742	5.240	< .001	-34.604	-8.879
		GP	-27.696	5.522	< .001	-41.250	-14.142

continued . . .

Table A.81: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
yes, $<$ 3 hpw	SS	GP	-4.868	4.795	.941	-16.639	6.903
		TS	22.405	5.240	$<$ .001	9.542	35.267
	GP	SS	4.868	4.795	.941	-6.903	16.639
		TS	27.272	5.522	$<$ .001	13.718	40.827
	TS	SS	-22.405	5.240	$<$ .001	-35.267	-9.542
		GP	-27.272	5.522	$<$ .001	-40.827	-13.718

Table A.82: Post-hoc pairwise comparisons based on observed means with different methods for PS in DRV scenario with VG frequency grouping factor. hpw - hours per week; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 274.506$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	no / h. ever	used to, $>$ 3 hpw	2.665	5.664	.990	-13.209	18.538
		used to, $<$ 3 hpw	.813	6.676	1.000	-17.894	19.520
		yes, $>$ 3 hpw	-11.720	6.381	.362	-29.603	6.163
		yes, $<$ 3 hpw	-10.487	6.381	.476	-28.370	7.396
	used to, $>$ 3 hpw	no / h. ever	-2.665	5.664	.990	-18.538	13.209
		used to, $<$ 3 hpw	-1.852	6.118	.998	-18.997	15.294
		yes, $>$ 3 hpw	-14.385	5.796	.107	-30.627	1.857
		yes, $<$ 3 hpw	-13.151	5.796	.168	-29.394	3.091
	used to, $<$ 3 hpw	no / h. ever	-.813	6.676	1.000	-19.520	17.894
		used to, $>$ 3 hpw	1.852	6.118	.998	-15.294	18.997
		yes, $>$ 3 hpw	-12.533	6.788	.356	-31.554	6.488
		yes, $<$ 3 hpw	-11.300	6.788	.462	-30.321	7.721
	yes, $>$ 3 hpw	no / h. ever	11.720	6.381	.362	-6.163	29.603
		used to, $>$ 3 hpw	14.385	5.796	.107	-1.857	30.627
		used to, $<$ 3 hpw	12.533	6.788	.356	-6.488	31.554
		yes, $<$ 3 hpw	1.233	6.499	1.000	-16.978	19.444

continued . . .

Table A.82: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	yes, < 3 hpw	no / h. ever	10.487	6.381	.476	-7.396	28.370
		used to, > 3 hpw	13.151	5.796	.168	-3.091	29.394
		used to, < 3 hpw	11.300	6.788	.462	-7.721	30.321
		yes, > 3 hpw	-1.233	6.499	1.000	-19.444	16.978
Bonferroni	no / h. ever	used to, > 3 hpw	2.665	5.664	1.000	-13.770	19.100
		used to, < 3 hpw	.813	6.676	1.000	-18.556	20.182
		yes, > 3 hpw	-11.720	6.381	.706	-30.236	6.796
		yes, < 3 hpw	-10.487	6.381	1.000	-29.003	8.029
	used to, > 3 hpw	no / h. ever	-2.665	5.664	1.000	-19.100	13.770
		used to, < 3 hpw	-1.852	6.118	1.000	-19.604	15.901
		yes, > 3 hpw	-14.385	5.796	.155	-31.202	2.432
		yes, < 3 hpw	-13.151	5.796	.264	-29.969	3.666
	used to, < 3 hpw	no / h. ever	-.813	6.676	1.000	-20.182	18.556
		used to, > 3 hpw	1.852	6.118	1.000	-15.901	19.604
		yes, > 3 hpw	-12.533	6.788	.692	-32.227	7.161
		yes, < 3 hpw	-11.300	6.788	1.000	-30.994	8.394
	yes, > 3 hpw	no / h. ever	11.720	6.381	.706	-6.796	30.236
		used to, > 3 hpw	14.385	5.796	.155	-2.432	31.202
		used to, < 3 hpw	12.533	6.788	.692	-7.161	32.227
		yes, < 3 hpw	1.233	6.499	1.000	-17.622	20.089
	yes, < 3 hpw	no / h. ever	10.487	6.381	1.000	-8.029	29.003
		used to, > 3 hpw	13.151	5.796	.264	-3.666	29.969
		used to, < 3 hpw	11.300	6.788	1.000	-8.394	30.994
		yes, > 3 hpw	-1.233	6.499	1.000	-20.089	17.622

ANOVA PS DRV with MG

Table A.83: Estimates for PS in DRV scenario with MG as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
MG usage	no / h. ever	67.70	3.171	61.38	74.03
	used to	59.53	3.818	51.91	67.14
	yes	61.06	3.485	54.11	68.01
Inc.	SS	65.53	2.610	60.32	70.73
	GP	73.41	2.388	68.65	78.17
	TS	49.36	2.150	45.07	53.64
no / h. ever	* SS	73.31	4.094	65.14	81.47
	* GP	79.46	3.746	71.99	86.93
	* TS	50.34	3.372	43.61	57.06
used to	* SS	59.45	4.930	49.62	69.29
	* GP	70.39	4.511	61.40	79.39
	* TS	48.74	4.060	40.64	56.83
yes	* SS	63.81	4.501	54.84	72.79
	* GP	70.38	4.118	62.16	78.59
	* TS	49.00	3.706	41.61	56.39

Table A.84: Pairwise comparisons of main effects of inceptor and MG, based on estimated marginal means for PS in DRV scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
MG usage	no / h. ever	used to	8.175	4.963	.312	-3.999	20.348
		yes	6.641	4.712	.489	-4.917	18.198
	used to	no / h. ever	-8.175	4.963	.312	-20.348	3.999
		yes	-1.534	5.170	1.000	-14.214	11.147
	yes	no / h. ever	-6.641	4.712	.489	-18.198	4.917

continued . . .

Table A.84: . . . continued

Factor			MD	SE	p	LB	UB
MG usage	yes	used to	1.534	5.170	1.000	-11.147	14.214
Inc.	SS	GP	-7.885	2.014	.001	-12.825	-2.946
		TS	16.169	2.229	$< .001$	10.701	21.637
	GP	SS	7.885	2.014	.001	2.946	12.825
		TS	24.054	2.370	$< .001$	18.241	29.867
	TS	SS	-16.169	2.229	$< .001$	-21.637	-10.701
		GP	-24.054	2.370	$< .001$	-29.867	-18.241

Table A.85: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with MG as a grouping factor showing patterns of interaction of inceptor and MG. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	no / h. ever	used to	13.854	6.409	.102	-1.865	29.574
		yes	9.496	6.084	.369	-5.428	24.420
	used to	no / h. ever	-13.854	6.409	.102	-29.574	1.865
		yes	-4.358	6.676	1.000	-20.732	12.016
	yes	no / h. ever	-9.496	6.084	.369	-24.420	5.428
		used to	4.358	6.676	1.000	-12.016	20.732
GP	no / h. ever	used to	9.070	5.864	.379	-5.314	23.453
		yes	9.088	5.567	.321	-4.568	22.743
	used to	no / h. ever	-9.070	5.864	.379	-23.453	5.314
		yes	.018	6.108	1.000	-14.964	15.001
	yes	no / h. ever	-9.088	5.567	.321	-22.743	4.568
		used to	-.018	6.108	1.000	-15.001	14.964
TS	no / h. ever	used to	1.601	5.277	1.000	-11.344	14.546
		yes	1.339	5.010	1.000	-10.951	13.629
	used to	no / h. ever	-1.601	5.277	1.000	-14.546	11.344
		yes	-.262	5.497	1.000	-13.746	13.222

continued . . .

Table A.85: . . . continued

Inc.		Grouping factor	MD	SE	p	LB	UB
TS	yes	no / h. ever	-1.339	5.010	1.000	-13.629	10.951
		used to	.262	5.497	1.000	-13.222	13.746

Table A.86: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with MG as a grouping factor showing patterns of interaction of MG and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
no / h. ever	SS	GP	-6.154	3.158	.166	-13.902	1.593
		TS	22.972	3.497	$< .001$	14.396	31.549
	GP	SS	6.154	3.158	.166	-1.593	13.902
		TS	29.127	3.717	$< .001$	20.009	38.244
	TS	SS	-22.972	3.497	$< .001$	-31.549	-14.396
		GP	-29.127	3.717	$< .001$	-38.244	-20.009
used to	SS	GP	-10.939	3.803	.016	-20.268	-1.610
		TS	10.719	4.210	.039	.391	21.046
	GP	SS	10.939	3.803	.016	1.610	20.268
		TS	21.658	4.476	$< .001$	10.679	32.637
	TS	SS	-10.719	4.210	.039	-21.046	-.391
		GP	-21.658	4.476	$< .001$	-32.637	-10.679
yes	SS	GP	-6.563	3.472	.189	-15.079	1.953
		TS	14.815	3.844	.001	5.387	24.243
	GP	SS	6.563	3.472	.189	-1.953	15.079
		TS	21.378	4.086	$< .001$	11.355	31.400
	TS	SS	-14.815	3.844	.001	-24.243	-5.387
		GP	-21.378	4.086	$< .001$	-31.400	-11.355

Table A.87: Post-hoc pairwise comparisons based on observed means with different methods for PS in DRV scenario with MG grouping factor. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 291.550$.

Method	Grouping factor		MD	SE	*p*	LB	UB
Tukey HSD	no / h. ever	used to	8.175	4.963	.233	-3.709	20.059
		yes	6.641	4.712	.342	-4.642	17.924
	used to	no / h. ever	-8.175	4.963	.233	-20.059	3.709
		yes	-1.534	5.170	.953	-13.913	10.845
	yes	no / h. ever	-6.641	4.712	.342	-17.924	4.642
		used to	1.534	5.170	.953	-10.845	13.913
Bonferroni	no / h. ever	used to	8.175	4.963	.312	-3.999	20.348
		yes	6.641	4.712	.489	-4.917	18.198
	used to	no / h. ever	-8.175	4.963	.312	-20.348	3.999
		yes	-1.534	5.170	1.000	-14.214	11.147
	yes	no / h. ever	-6.641	4.712	.489	-18.198	4.917
		used to	1.534	5.170	1.000	-11.147	14.214

ANOVA PS DRV with inceptor order

Table A.88: Estimates for PS in DRV scenario with inceptor order as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Inc. order	123	61.13	4.875	51.40	70.86
	132	58.22	5.074	48.09	68.35
	213	62.40	4.875	52.67	72.13
	231	66.29	5.300	55.71	76.87
	312	66.76	5.300	56.18	77.34
	321	65.49	4.875	55.76	75.22

continued . . .

Table A.88: . . . continued

Effect		M	SE	LB	UB
	SS	66.63	2.636	61.37	71.89
Inc.	GP	74.09	2.446	69.21	78.97
	TS	49.42	2.079	45.27	53.57
	* SS	56.86	6.229	44.43	69.29
123	* GP	74.28	5.780	62.74	85.82
	* TS	52.25	4.913	42.45	62.06
	* SS	61.24	6.484	48.30	74.18
132	* GP	74.65	6.016	62.64	86.66
	* TS	38.77	5.114	28.56	48.97
	* SS	64.96	6.229	52.53	77.40
213	* GP	67.45	5.780	55.92	78.99
	* TS	54.79	4.913	44.98	64.60
	* SS	71.41	6.772	57.90	84.93
231	* GP	73.57	6.284	61.03	86.11
	* TS	53.88	5.341	43.22	64.54
	* SS	73.33	6.772	59.82	86.85
312	* GP	78.05	6.284	65.51	90.59
	* TS	48.90	5.341	38.24	59.56
	* SS	71.98	6.229	59.54	84.41
321	* GP	76.55	5.780	65.02	88.09
	* TS	47.93	4.913	38.12	57.73

Table A.89: Pairwise comparisons of main effects of inceptor and inceptor order, based on estimated marginal means for PS in DRV scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Inc. order	123	132	2.913	7.037	1.000	-18.508	24.333

continued . . .

Table A.89: ... continued

Factor			MD	SE	p	LB	UB
Inc. order	123	213	-1.273	6.895	1.000	-22.261	19.715
		231	-5.158	7.201	1.000	-27.079	16.763
		312	-5.630	7.201	1.000	-27.551	16.291
		321	-4.356	6.895	1.000	-25.344	16.632
	132	123	-2.913	7.037	1.000	-24.333	18.508
		213	-4.185	7.037	1.000	-25.606	17.235
		231	-8.071	7.337	1.000	-30.407	14.265
		312	-8.542	7.337	1.000	-30.878	13.793
		321	-7.269	7.037	1.000	-28.689	14.152
	213	123	1.273	6.895	1.000	-19.715	22.261
		132	4.185	7.037	1.000	-17.235	25.606
		231	-3.886	7.201	1.000	-25.807	18.035
		312	-4.357	7.201	1.000	-26.278	17.564
		321	-3.083	6.895	1.000	-24.071	17.905
	231	123	5.158	7.201	1.000	-16.763	27.079
		132	8.071	7.337	1.000	-14.265	30.407
		213	3.886	7.201	1.000	-18.035	25.807
		312	-.471	7.495	1.000	-23.287	22.345
		321	.802	7.201	1.000	-21.119	22.723
	312	123	5.630	7.201	1.000	-16.291	27.551
		132	8.542	7.337	1.000	-13.793	30.878
		213	4.357	7.201	1.000	-17.564	26.278
		231	.471	7.495	1.000	-22.345	23.287
		321	1.274	7.201	1.000	-20.647	23.195
	321	123	4.356	6.895	1.000	-16.632	25.344
		132	7.269	7.037	1.000	-14.152	28.689
		213	3.083	6.895	1.000	-17.905	24.071
		231	-.802	7.201	1.000	-22.723	21.119
		312	-1.274	7.201	1.000	-23.195	20.647

continued ...

Table A.89: . . . continued

Factor			MD	SE	p	LB	UB
Inc.	SS	GP	-7.460	1.925	.001	-12.188	-2.733
		TS	17.212	2.153	$<$ .001	11.926	22.498
	GP	SS	7.460	1.925	.001	2.733	12.188
		TS	24.672	2.262	$<$ .001	19.117	30.228
	TS	SS	-17.212	2.153	$<$ .001	-22.498	-11.926
		GP	-24.672	2.262	$<$ .001	-30.228	-19.117

Table A.90: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with inceptor order as a grouping factor showing patterns of interaction of inceptor and inceptor order. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	123	132	-4.378	8.991	1.000	-31.748	22.993
		213	-8.104	8.810	1.000	-34.922	18.714
		231	-14.554	9.202	1.000	-42.564	13.456
		312	-16.474	9.202	1.000	-44.484	11.536
		321	-15.117	8.810	1.000	-41.935	11.701
	132	123	4.378	8.991	1.000	-22.993	31.748
		213	-3.726	8.991	1.000	-31.097	23.644
		231	-10.176	9.376	1.000	-38.716	18.364
		312	-12.096	9.376	1.000	-40.636	16.444
		321	-10.739	8.991	1.000	-38.110	16.631
	213	123	8.104	8.810	1.000	-18.714	34.922
		132	3.726	8.991	1.000	-23.644	31.097
		231	-6.450	9.202	1.000	-34.460	21.560
		312	-8.370	9.202	1.000	-36.380	19.640
		321	-7.013	8.810	1.000	-33.831	19.805
	231	123	14.554	9.202	1.000	-13.456	42.564
		132	10.176	9.376	1.000	-18.364	38.716

continued . . .

Table A.90: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	231	213	6.450	9.202	1.000	-21.560	34.460
		312	-1.920	9.577	1.000	-31.074	27.234
		321	-.563	9.202	1.000	-28.573	27.447
	312	123	16.474	9.202	1.000	-11.536	44.484
		132	12.096	9.376	1.000	-16.444	40.636
		213	8.370	9.202	1.000	-19.640	36.380
		231	1.920	9.577	1.000	-27.234	31.074
		321	1.357	9.202	1.000	-26.653	29.367
	321	123	15.117	8.810	1.000	-11.701	41.935
		132	10.739	8.991	1.000	-16.631	38.110
		213	7.013	8.810	1.000	-19.805	33.831
		231	.563	9.202	1.000	-27.447	28.573
		312	-1.357	9.202	1.000	-29.367	26.653
GP	123	132	-.368	8.343	1.000	-25.765	25.028
		213	6.825	8.174	1.000	-18.058	31.709
		231	.709	8.538	1.000	-25.281	26.700
		312	-3.768	8.538	1.000	-29.759	22.222
		321	-2.274	8.174	1.000	-27.157	22.610
	132	123	.368	8.343	1.000	-25.028	25.765
		213	7.194	8.343	1.000	-18.203	32.591
		231	1.078	8.699	1.000	-25.404	27.560
		312	-3.400	8.699	1.000	-29.882	23.082
		321	-1.905	8.343	1.000	-27.302	23.492
	213	123	-6.825	8.174	1.000	-31.709	18.058
		132	-7.194	8.343	1.000	-32.591	18.203
		231	-6.116	8.538	1.000	-32.106	19.874
		312	-10.594	8.538	1.000	-36.584	15.397
		321	-9.099	8.174	1.000	-33.983	15.785
	231	123	-.709	8.538	1.000	-26.700	25.281

continued . . .

Table A.90: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
GP	231	132	-1.078	8.699	1.000	-27.560	25.404
		213	6.116	8.538	1.000	-19.874	32.106
		312	-4.478	8.887	1.000	-31.529	22.574
		321	-2.983	8.538	1.000	-28.973	23.007
	312	123	3.768	8.538	1.000	-22.222	29.759
		132	3.400	8.699	1.000	-23.082	29.882
		213	10.594	8.538	1.000	-15.397	36.584
		231	4.478	8.887	1.000	-22.574	31.529
		321	1.495	8.538	1.000	-24.496	27.485
	321	123	2.274	8.174	1.000	-22.610	27.157
		132	1.905	8.343	1.000	-23.492	27.302
		213	9.099	8.174	1.000	-15.785	33.983
		231	2.983	8.538	1.000	-23.007	28.973
		312	-1.495	8.538	1.000	-27.485	24.496
TS	123	132	13.484	7.091	.923	-8.102	35.070
		213	-2.540	6.948	1.000	-23.690	18.610
		231	-1.631	7.257	1.000	-23.721	20.460
		312	3.353	7.257	1.000	-18.738	25.444
		321	4.323	6.948	1.000	-16.827	25.473
	132	123	-13.484	7.091	.923	-35.070	8.102
		213	-16.024	7.091	.406	-37.610	5.562
		231	-15.115	7.394	.673	-37.623	7.394
		312	-10.131	7.394	1.000	-32.640	12.378
		321	-9.161	7.091	1.000	-30.748	12.425
	213	123	2.540	6.948	1.000	-18.610	23.690
		132	16.024	7.091	.406	-5.562	37.610
		231	.909	7.257	1.000	-21.181	23< .001
		312	5.893	7.257	1.000	-16.198	27.984
		321	6.863	6.948	1.000	-14.288	28.013

continued . . .

Table A.90: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
TS	231	123	1.631	7.257	1.000	-20.460	23.721
		132	15.115	7.394	.673	-7.394	37.623
		213	-.909	7.257	1.000	-23< .001	21.181
		312	4.984	7.553	1.000	-18.009	27.976
		321	5.953	7.257	1.000	-16.137	28.044
	312	123	-3.353	7.257	1.000	-25.444	18.738
		132	10.131	7.394	1.000	-12.378	32.640
		213	-5.893	7.257	1.000	-27.984	16.198
		231	-4.984	7.553	1.000	-27.976	18.009
		321	.970	7.257	1.000	-21.121	23.060
	321	123	-4.323	6.948	1.000	-25.473	16.827
		132	9.161	7.091	1.000	-12.425	30.748
		213	-6.863	6.948	1.000	-28.013	14.288
		231	-5.953	7.257	1.000	-28.044	16.137
		312	-.970	7.257	1.000	-23.060	21.121

Table A.91: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with inceptor order as a grouping factor showing patterns of interaction of inceptor order and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
123	SS	GP	-17.419	4.549	.001	-28.589	-6.248
		TS	4.609	5.087	1.000	-7.882	17.101
	GP	SS	17.419	4.549	.001	6.248	28.589
		TS	22.028	5.346	$< .001$	8.900	35.156
	TS	SS	-4.609	5.087	1.000	-17.101	7.882
		GP	-22.028	5.346	$< .001$	-35.156	-8.900
132	SS	GP	-13.410	4.735	.018	-25.036	-1.783
		TS	22.471	5.295	$< .001$	9.470	35.472

continued . . .

Table A.91: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
132	GP	SS	13.410	4.735	.018	1.783	25.036
		TS	35.881	5.565	< .001	22.216	49.545
	TS	SS	-22.471	5.295	< .001	-35.472	-9.470
		GP	-35.881	5.565	< .001	-49.545	-22.216
213	SS	GP	-2.490	4.549	1.000	-13.660	8.681
		TS	10.173	5.087	.149	-2.318	22.665
	GP	SS	2.490	4.549	1.000	-8.681	13.660
		TS	12.663	5.346	.062	-.465	25.791
	TS	SS	-10.173	5.087	.149	-22.665	2.318
		GP	-12.663	5.346	.062	-25.791	.465
231	SS	GP	-2.156	4.945	1.000	-14.300	9.988
		TS	17.533	5.530	.007	3.953	31.112
	GP	SS	2.156	4.945	1.000	-9.988	14.300
		TS	19.688	5.812	.004	5.416	33.960
	TS	SS	-17.533	5.530	.007	-31.112	-3.953
		GP	-19.688	5.812	.004	-33.960	-5.416
312	SS	GP	-4.713	4.945	1.000	-16.857	7.430
		TS	24.436	5.530	< .001	10.857	38.016
	GP	SS	4.713	4.945	1.000	-7.430	16.857
		TS	29.150	5.812	< .001	14.878	43.422
	TS	SS	-24.436	5.530	< .001	-38.016	-10.857
		GP	-29.150	5.812	< .001	-43.422	-14.878
321	SS	GP	-4.576	4.549	.954	-15.746	6.595
		TS	24.049	5.087	< .001	11.558	36.540
	GP	SS	4.576	4.549	.954	-6.595	15.746
		TS	28.625	5.346	< .001	15.497	41.753
	TS	SS	-24.049	5.087	< .001	-36.540	-11.558
		GP	-28.625	5.346	< .001	-41.753	-15.497

Table A.92: Post-hoc pairwise comparisons based on observed means with different methods for PS in DRV scenario with inceptor order grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 308.981$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	123	132	2.913	7.037	.998	-17.732	23.557
		213	-1.273	6.895	1.000	-21.500	18.955
		231	-5.158	7.201	.979	-26.285	15.968
		312	-5.630	7.201	.970	-26.756	15.497
		321	-4.356	6.895	.988	-24.583	15.871
	132	123	-2.913	7.037	.998	-23.557	17.732
		213	-4.185	7.037	.991	-24.830	16.459
		231	-8.071	7.337	.880	-29.598	13.455
		312	-8.542	7.337	.852	-30.069	12.984
		321	-7.269	7.037	.905	-27.913	13.376
	213	123	1.273	6.895	1.000	-18.955	21.500
		132	4.185	7.037	.991	-16.459	24.830
		231	-3.886	7.201	.994	-25.012	17.241
		312	-4.357	7.201	.990	-25.484	16.770
		321	-3.083	6.895	.998	-23.311	17.144
	231	123	5.158	7.201	.979	-15.968	26.285
		132	8.071	7.337	.880	-13.455	29.598
		213	3.886	7.201	.994	-17.241	25.012
		312	-.471	7.495	1.000	-22.461	21.518
		321	.802	7.201	1.000	-20.324	21.929
	312	123	5.630	7.201	.970	-15.497	26.756
		132	8.542	7.337	.852	-12.984	30.069
		213	4.357	7.201	.990	-16.770	25.484
		231	.471	7.495	1.000	-21.518	22.461
		321	1.274	7.201	1.000	-19.853	22.401
	321	123	4.356	6.895	.988	-15.871	24.583

continued ...

Table A.92: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	321	132	7.269	7.037	.905	-13.376	27.913
		213	3.083	6.895	.998	-17.144	23.311
		231	-.802	7.201	1.000	-21.929	20.324
		312	-1.274	7.201	1.000	-22.401	19.853
Bonferroni	123	132	2.913	7.037	1.000	-18.508	24.333
		213	-1.273	6.895	1.000	-22.261	19.715
		231	-5.158	7.201	1.000	-27.079	16.763
		312	-5.630	7.201	1.000	-27.551	16.291
		321	-4.356	6.895	1.000	-25.344	16.632
	132	123	-2.913	7.037	1.000	-24.333	18.508
		213	-4.185	7.037	1.000	-25.606	17.235
		231	-8.071	7.337	1.000	-30.407	14.265
		312	-8.542	7.337	1.000	-30.878	13.793
		321	-7.269	7.037	1.000	-28.689	14.152
	213	123	1.273	6.895	1.000	-19.715	22.261
		132	4.185	7.037	1.000	-17.235	25.606
		231	-3.886	7.201	1.000	-25.807	18.035
		312	-4.357	7.201	1.000	-26.278	17.564
		321	-3.083	6.895	1.000	-24.071	17.905
	231	123	5.158	7.201	1.000	-16.763	27.079
		132	8.071	7.337	1.000	-14.265	30.407
		213	3.886	7.201	1.000	-18.035	25.807
		312	-.471	7.495	1.000	-23.287	22.345
		321	.802	7.201	1.000	-21.119	22.723
	312	123	5.630	7.201	1.000	-16.291	27.551
		132	8.542	7.337	1.000	-13.793	30.878
		213	4.357	7.201	1.000	-17.564	26.278
		231	.471	7.495	1.000	-22.345	23.287
		321	1.274	7.201	1.000	-20.647	23.195

continued . . .

Table A.92: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Bonferroni	321	123	4.356	6.895	1.000	-16.632	25.344
		132	7.269	7.037	1.000	-14.152	28.689
		213	3.083	6.895	1.000	-17.905	24.071
		231	-.802	7.201	1.000	-22.723	21.119
		312	-1.274	7.201	1.000	-23.195	20.647

ANOVA PS DRV with FE group

Table A.93: Estimates for PS in DRV scenario with FE group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
FE gr.	A	71.67	3.387	64.92	78.43
	B	71.10	3.148	64.82	77.38
	C	52.92	2.610	47.71	58.13
Inc.	SS	68.71	2.371	63.98	73.44
	GP	76.48	2.062	72.37	80.59
	TS	50.51	2.050	46.42	54.60
A	* SS	75.19	4.537	66.14	84.24
	* GP	88.49	3.947	80.61	96.36
	* TS	51.34	3.924	43.51	59.16
B	* SS	77.19	4.217	68.78	85.60
	* GP	79.03	3.668	71.72	86.35
	* TS	57.08	3.647	49.80	64.35
C	* SS	53.74	3.496	46.77	60.71
	* GP	61.92	3.041	55.85	67.98
	* TS	43.10	3.024	37.07	49.13

Table A.94: Pairwise comparisons of main effects of inceptor and FE group, based on estimated marginal means for PS in DRV scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
FE gr.	A	B	.571	4.624	1.000	-10.771	11.912
		C	18.751	4.276	< .001	8.263	29.240
	B	A	-.571	4.624	1.000	-11.912	10.771
		C	18.181	4.089	< .001	8.151	28.210
	C	A	-18.751	4.276	< .001	-29.240	-8.263
		B	-18.181	4.089	< .001	-28.210	-8.151
Inc.	SS	GP	-7.771	1.985	.001	-12.641	-2.901
		TS	18.202	2.236	< .001	12.717	23.686
	GP	SS	7.771	1.985	.001	2.901	12.641
		TS	25.972	2.259	< .001	20.432	31.513
	TS	SS	-18.202	2.236	< .001	-23.686	-12.717
		GP	-25.972	2.259	< .001	-31.513	-20.432

Table A.95: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with FE group as a grouping factor showing patterns of interaction of inceptor and FE group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	A	B	-2.002	6.194	1.000	-17.195	13.191
		C	21.451	5.728	.001	7.401	35.501
	B	A	2.002	6.194	1.000	-13.191	17.195
		C	23.453	5.477	< .001	10.017	36.888
	C	A	-21.451	5.728	.001	-35.501	-7.401
		B	-23.453	5.477	< .001	-36.888	-10.017
GP	A	B	9.453	5.388	.251	-3.763	22.668
		C	26.567	4.982	< .001	14.346	38.788

continued . . .

Table A.95: ... continued

Inc.	Grouping factor		MD	SE	p	LB	UB
GP	B	A	-9.453	5.388	.251	-22.668	3.763
		C	17.114	4.764	.002	5.428	28.801
	C	A	-26.567	4.982	< .001	-38.788	-14.346
		B	-17.114	4.764	.002	-28.801	-5.428
TS	A	B	-5.738	5.357	.863	-18.878	7.401
		C	8.236	4.954	.303	-3.915	20.387
	B	A	5.738	5.357	.863	-7.401	18.878
		C	13.975	4.737	.013	2.355	25.594
	C	A	-8.236	4.954	.303	-20.387	3.915
		B	-13.975	4.737	.013	-25.594	-2.355

Table A.96: Pairwise comparisons based on estimated marginal means for PS in DRV scenario with FE group as a grouping factor showing patterns of interaction of FE group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
A	SS	GP	-13.295	3.800	.002	-22.615	-3.975
		TS	23.852	4.279	< .001	13.355	34.348
	GP	SS	13.295	3.800	.002	3.975	22.615
		TS	37.146	4.322	< .001	26.544	47.749
	TS	SS	-23.852	4.279	< .001	-34.348	-13.355
		GP	-37.146	4.322	< .001	-47.749	-26.544
B	SS	GP	-1.840	3.531	1.000	-10.501	6.822
		TS	20.116	3.977	< .001	10.361	29.870
	GP	SS	1.840	3.531	1.000	-6.822	10.501
		TS	21.955	4.017	< .001	12.102	31.808
	TS	SS	-20.116	3.977	< .001	-29.870	-10.361
		GP	-21.955	4.017	< .001	-31.808	-12.102

continued ...

Table A.96: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
C	SS	GP	-8.178	2.928	.020	-15.360	-.996
		TS	10.638	3.297	.006	2.549	18.726
	GP	SS	8.178	2.928	.020	.996	15.360
		TS	18.815	3.331	$< .001$	10.646	26.985
	TS	SS	-10.638	3.297	.006	-18.726	-2.549
		GP	-18.815	3.331	$< .001$	-26.985	-10.646

Table A.97: Post-hoc pairwise comparisons based on observed means with different methods for PS in DRV scenario with FE group grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 217.963$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	A	B	.571	4.624	.992	-10.501	11.643
		C	18.751	4.276	$< .001$	8.513	28.990
	B	A	-.571	4.624	.992	-11.643	10.501
		C	18.181	4.089	$< .001$	8.390	27.972
	C	A	-18.751	4.276	$< .001$	-28.990	-8.513
		B	-18.181	4.089	$< .001$	-27.972	-8.390
Bonferroni	A	B	.571	4.624	1.000	-10.771	11.912
		C	18.751	4.276	$< .001$	8.263	29.240
	B	A	-.571	4.624	1.000	-11.912	10.771
		C	18.181	4.089	$< .001$	8.151	28.210
	C	A	-18.751	4.276	$< .001$	-29.240	-8.263
		B	-18.181	4.089	$< .001$	-28.210	-8.151

ANOVA PS DRH with gender

Table A.98: Descriptive statistics for PS in DRH scenario with gender as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Female	67.40	27.09	82.02	22.97	63.16	24.43	18
Male	78.79	21.63	90.27	14.90	73.93	20.44	54
Pref. not to say	87.00		97.81		82.26		1
Total	76.10	23.32	88.34	17.38	71.39	21.72	73

Table A.99: Estimates for PS in DRH scenario with gender as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Gender	Female	70.86	4.256	62.37	79.35
	Male	81.00	2.457	76.09	85.90
	Pref. not to say	89.02	18.059	53.01	125.04
Inc.	SS	77.73	7.972	61.83	93.63
	GP	90.03	5.946	78.17	101.89
	TS	73.11	7.419	58.32	87.91
Female	* SS	67.40	5.440	56.55	78.24
	* GP	82.02	4.057	73.93	90.11
	* TS	63.16	5.062	53.06	73.25
Male	* SS	78.79	3.141	72.53	85.06
	* GP	90.27	2.342	85.60	94.94
	* TS	73.93	2.923	68.10	79.76
Pref. not to say	* SS	87.00	23.078	40.97	133.03
	* GP	97.81	17.211	63.49	132.14
	* TS	82.26	21.476	39.43	125.09

Table A.100: Pairwise comparisons of main effects of inceptor and gender, based on estimated marginal means for PS in DRH scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Gender	Female	Male	-10.139	4.915	.128	-22.195	1.917
		Pref. not to say	-18.168	18.554	.993	-63.677	27.342
	Male	Female	10.139	4.915	.128	-1.917	22.195
		Pref. not to say	-8.028	18.225	1.000	-52.733	36.676
	Pref. not to say	Female	18.168	18.554	.993	-27.342	63.677
		Male	8.028	18.225	1.000	-36.676	52.733
Inc.	SS	GP	-12.303	6.517	.190	-28.288	3.682
		TS	4.615	6.216	1.000	-10.631	19.862
	GP	SS	12.303	6.517	.190	-3.682	28.288
		TS	16.919	5.521	.009	3.376	30.461
	TS	SS	-4.615	6.216	1.000	-19.862	10.631
		GP	-16.919	5.521	.009	-30.461	-3.376

Table A.101: Pairwise comparisons based on estimated marginal means for PS in DRH scenario with gender as a grouping factor showing patterns of interaction of inceptor and gender. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	Female	Male	-11.398	6.281	.222	-26.804	4.009
		Pref. not to say	-19.604	23.710	1.000	-77.763	38.554
	Male	Female	11.398	6.281	.222	-4.009	26.804
		Pref. not to say	-8.206	23.291	1.000	-65.335	48.922
	Pref. not to say	Female	19.604	23.710	1.000	-38.554	77.763
		Male	8.206	23.291	1.000	-48.922	65.335
GP	Female	Male	-8.251	4.684	.248	-19.741	3.239
		Pref. not to say	-15.798	17.683	1.000	-59.171	27.575

continued . . .

Table A.101: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
GP	Male	Female	8.251	4.684	.248	-3.239	19.741
		Pref. not to say	-7.547	17.370	1.000	-50.153	35.058
	Pref. not to say	Female	15.798	17.683	1.000	-27.575	59.171
		Male	7.547	17.370	1.000	-35.058	50.153
TS	Female	Male	-10.769	5.845	.209	-25.106	3.568
		Pref. not to say	-19.100	22.064	1.000	-73.222	35.021
	Male	Female	10.769	5.845	.209	-3.568	25.106
		Pref. not to say	-8.332	21.674	1.000	-61.495	44.832
	Pref. not to say	Female	19.100	22.064	1.000	-35.021	73.222
		Male	8.332	21.674	1.000	-44.832	61.495

Table A.102: Pairwise comparisons based on estimated marginal means for PS in DRH scenario with gender as a grouping factor showing patterns of interaction of gender and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
Female	SS	GP	-14.621	4.446	.005	-25.528	-3.715
		TS	4.237	4.241	.963	-6.165	14.640
	GP	SS	14.621	4.446	.005	3.715	25.528
		TS	18.859	3.767	< .001	9.619	28.099
	TS	SS	-4.237	4.241	.963	-14.640	6.165
		GP	-18.859	3.767	< .001	-28.099	-9.619
Male	SS	GP	-11.474	2.567	< .001	-17.771	-5.177
		TS	4.867	2.448	.152	-1.139	10.873
	GP	SS	11.474	2.567	< .001	5.177	17.771
		TS	16.341	2.175	< .001	11.006	21.676
	TS	SS	-4.867	2.448	.152	-10.873	1.139
		GP	-16.341	2.175	< .001	-21.676	-11.006
Pref. not to say	SS	GP	-10.815	18.864	1.000	-57.087	35.457

continued . . .

Table A.102: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
Pref. not to say	SS	TS	4.741	17.993	1.000	-39.392	48.875
	GP	SS	10.815	18.864	1.000	-35.457	57.087
		TS	15.556	15.982	1.000	-23.645	54.758
	TS	SS	-4.741	17.993	1.000	-48.875	39.392
		GP	-15.556	15.982	1.000	-54.758	23.645

ANOVA PS DRH with FE group

Table A.103: Estimates for PS in DRH scenario with FE group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
FE gr.	A	87.94	4.047	79.87	96.01
	B	78.22	3.761	70.72	85.72
	C	73.33	3.118	67.11	79.55
Inc.	SS	77.79	2.663	72.48	83.11
	GP	89.20	2.049	85.12	93.29
	TS	72.49	2.548	67.41	77.57
A	* SS	88.10	5.097	77.93	98.26
	* GP	95.04	3.921	87.22	102.86
	* TS	80.68	4.877	70.95	90.41
B	* SS	76.81	4.737	67.36	86.26
	* GP	87.96	3.644	80.69	95.22
	* TS	69.89	4.532	60.85	78.93
C	* SS	68.48	3.928	60.65	76.31
	* GP	84.62	3.021	78.59	90.64
	* TS	66.90	3.758	59.40	74.39

Table A.104: Pairwise comparisons of main effects of inceptor and FE group, based on estimated marginal means for PS in DRH scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
FE gr.	A	B	9.722	5.524	.248	-3.828	23.272
		C	14.606	5.108	.017	2.076	27.137
	B	A	-9.722	5.524	.248	-23.272	3.828
		C	4.885	4.885	.962	-7.098	16.867
	C	A	-14.606	5.108	.017	-27.137	-2.076
		B	-4.885	4.885	.962	-16.867	7.098
Inc.	SS	GP	-11.410	2.220	$< .001$	-16.855	-5.965
		TS	5.307	2.130	.045	.081	10.532
	GP	SS	11.410	2.220	$< .001$	5.965	16.855
		TS	16.717	1.911	$< .001$	12.029	21.405
	TS	SS	-5.307	2.130	.045	-10.532	-.081
		GP	-16.717	1.911	$< .001$	-21.405	-12.029

Table A.105: Pairwise comparisons based on estimated marginal means for PS in DRH scenario with FE group as a grouping factor showing patterns of interaction of inceptor and FE group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	A	B	11.287	6.958	.328	-5.781	28.354
		C	19.615	6.435	.010	3.832	35.398
	B	A	-11.287	6.958	.328	-28.354	5.781
		C	8.328	6.153	.541	-6.765	23.421
	C	A	-19.615	6.435	.010	-35.398	-3.832
		B	-8.328	6.153	.541	-23.421	6.765
GP	A	B	7.084	5.352	.570	-6.045	20.213
		C	10.423	4.950	.116	-1.718	22.564

continued ...

Table A.105: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
GP	B	A	-7.084	5.352	.570	-20.213	6.045
		C	3.339	4.733	1.000	-8.271	14.949
	C	A	-10.423	4.950	.116	-22.564	1.718
		B	-3.339	4.733	1.000	-14.949	8.271
TS	A	B	10.794	6.658	.328	-5.537	27.125
		C	13.781	6.157	.085	-1.322	28.883
	B	A	-10.794	6.658	.328	-27.125	5.537
		C	2.987	5.888	1.000	-11.455	17.429
	C	A	-13.781	6.157	.085	-28.883	1.322
		B	-2.987	5.888	1.000	-17.429	11.455

Table A.106: Pairwise comparisons based on estimated marginal means for PS in DRH scenario with FE group as a grouping factor showing patterns of interaction of FE group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
A	SS	GP	-6.945	4.248	.320	-17.365	3.475
		TS	7.416	4.077	.220	-2.584	17.416
	GP	SS	6.945	4.248	.320	-3.475	17.365
		TS	14.361	3.658	.001	5.390	23.333
	TS	SS	-7.416	4.077	.220	-17.416	2.584
		GP	-14.361	3.658	.001	-23.333	-5.390
B	SS	GP	-11.148	3.948	.019	-20.831	-1.464
		TS	6.923	3.789	.216	-2.370	16.217
	GP	SS	11.148	3.948	.019	1.464	20.831
		TS	18.071	3.399	$<.001$	9.733	26.408
	TS	SS	-6.923	3.789	.216	-16.217	2.370
		GP	-18.071	3.399	$<.001$	-26.408	-9.733
C	SS	GP	-16.137	3.273	$<.001$	-24.166	-8.108

continued . . .

Table A.106: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
C	SS	TS	1.582	3.142	1.000	-6.124	9.287
	GP	SS	16.137	3.273	$< .001$	8.108	24.166
		TS	17.719	2.818	$< .001$	10.806	24.632
	TS	SS	-1.582	3.142	1.000	-9.287	6.124
		GP	-17.719	2.818	$< .001$	-24.632	-10.806

Table A.107: Post-hoc pairwise comparisons based on observed means with different methods for PS in DRH scenario with FE group grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 311.114$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	A	B	9.722	5.524	.191	-3.506	22.950
		C	14.606	5.108	.015	2.374	26.839
	B	A	-9.722	5.524	.191	-22.950	3.506
		C	4.885	4.885	.579	-6.813	16.582
	C	A	-14.606	5.108	.015	-26.839	-2.374
		B	-4.885	4.885	.579	-16.582	6.813
Bonferroni	A	B	9.722	5.524	.248	-3.828	23.272
		C	14.606	5.108	.017	2.076	27.137
	B	A	-9.722	5.524	.248	-23.272	3.828
		C	4.885	4.885	.962	-7.098	16.867
	C	A	-14.606	5.108	.017	-27.137	-2.076
		B	-4.885	4.885	.962	-16.867	7.098

ANOVA PS LN with gender

Table A.108: Descriptive statistics for PS in LN scenario with gender as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Female	67.18	31.97	58.19	33.79	57.57	20.52	18
Male	75.52	21.17	76.55	14.23	61.60	20.42	54
Pref. not to say	85.90		78.71		48.56		1
Total	73.60	24.22	72.05	21.96	60.43	20.29	73

Table A.109: Estimates for PS in LN scenario with gender as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Gender	Female	60.98	4.339	52.33	69.63
	Male	71.22	2.505	66.22	76.22
	Pref. not to say	71.06	18.409	34.34	107.77
Inc.	SS	76.20	8.374	59.50	92.90
	GP	71.15	7.168	56.85	85.45
	TS	55.91	7.064	41.82	70.00
Female	* SS	67.18	5.713	55.78	78.57
	* GP	58.19	4.891	48.44	67.95
	* TS	57.57	4.819	47.96	67.18
Male	* SS	75.52	3.298	68.94	82.10
	* GP	76.55	2.824	70.92	82.18
	* TS	61.60	2.783	56.05	67.15
Pref. not to say	* SS	85.90	24.239	37.56	134.24
	* GP	78.71	20.750	37.32	120.09
	* TS	48.56	20.447	7.78	89.34

Table A.110: Pairwise comparisons of main effects of inceptor and gender, based on estimated marginal means for PS in LN scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Gender	Female	Male	-10.242	5.010	.134	-22.531	2.048
		Pref. not to say	-10.078	18.913	1.000	-56.470	36.314
	Male	Female	10.242	5.010	.134	-2.048	22.531
		Pref. not to say	.164	18.578	1.000	-45.407	45.734
	Pref. not to say	Female	10.078	18.913	1.000	-36.314	56.470
		Male	-.164	18.578	1.000	-45.734	45.407
Inc.	SS	GP	5.049	6.964	1.000	-12.032	22.129
		TS	20.288	7.721	.032	1.350	39.227
	GP	SS	-5.049	6.964	1.000	-22.129	12.032
		TS	15.240	6.488	.065	-.674	31.153
	TS	SS	-20.288	7.721	.032	-39.227	-1.350
		GP	-15.240	6.488	.065	-31.153	.674

Table A.111: Pairwise comparisons based on estimated marginal means for PS in LN scenario with gender as a grouping factor showing patterns of interaction of inceptor and gender. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	Female	Male	-8.340	6.597	.631	-24.522	7.841
		Pref. not to say	-18.723	24.903	1.000	-79.807	42.361
	Male	Female	8.340	6.597	.631	-7.841	24.522
		Pref. not to say	-10.383	24.462	1.000	-70.386	49.620
	Pref. not to say	Female	18.723	24.903	1.000	-42.361	79.807
		Male	10.383	24.462	1.000	-49.620	70.386
GP	Female	Male	-18.357	5.647	.005	-32.209	-4.505
		Pref. not to say	-20.516	21.318	1.000	-72.807	31.774

continued ...

Table A.111: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
GP	Male	Female	18.357	5.647	.005	4.505	32.209
		Pref. not to say	-2.159	20.941	1.000	-53.524	49.206
	Pref. not to say	Female	20.516	21.318	1.000	-31.774	72.807
		Male	2.159	20.941	1.000	-49.206	53.524
TS	Female	Male	-4.028	5.565	1.000	-17.678	9.622
		Pref. not to say	9.006	21.008	1.000	-42.523	60.535
	Male	Female	4.028	5.565	1.000	-9.622	17.678
		Pref. not to say	13.034	20.636	1.000	-37.582	63.651
	Pref. not to say	Female	-9.006	21.008	1.000	-60.535	42.523
		Male	-13.034	20.636	1.000	-63.651	37.582

Table A.112: Pairwise comparisons based on estimated marginal means for PS in LN scenario with gender as a grouping factor showing patterns of interaction of gender and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
Female	SS	GP	8.985	4.751	.188	-2.669	20.639
		TS	9.608	5.268	.217	-3.313	22.529
	GP	SS	-8.985	4.751	.188	-20.639	2.669
		TS	.623	4.426	1.000	-10.235	11.480
	TS	SS	-9.608	5.268	.217	-22.529	3.313
		GP	-.623	4.426	1.000	-11.480	10.235
Male	SS	GP	-1.032	2.743	1.000	-7.760	5.697
		TS	13.920	3.041	$< .001$	6.460	21.380
	GP	SS	1.032	2.743	1.000	-5.697	7.760
		TS	14.952	2.556	$< .001$	8.683	21.220
	TS	SS	-13.920	3.041	$< .001$	-21.380	-6.460
		GP	-14.952	2.556	$< .001$	-21.220	-8.683
Pref. not to say	SS	GP	7.192	20.158	1.000	-42.252	56.636

continued . . .

Table A.112: . . . continued

Subgroup	Inc.		MD	SE	*p*	LB	UB
Pref. not to say	SS	TS	37.337	22.349	.298	-17.483	92.158
	GP	SS	-7.192	20.158	1.000	-56.636	42.252
		TS	30.145	18.780	.339	-15.920	76.210
	TS	SS	-37.337	22.349	.298	-92.158	17.483
		GP	-30.145	18.780	.339	-76.210	15.920

ANOVA PS LN with hand

Table A.113: Descriptive statistics for PS in LN scenario with handedness as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Ambidextrous	94.40	6.08	87.65	7.90	79.41	3.42	3
Left-handed	50.67	28.79	72.46	13.92	50.97	28.10	7
Right-handed	75.16	22.73	71.26	22.96	60.57	19.33	63
Total	73.60	24.22	72.05	21.96	60.43	20.29	73

Table A.114: Estimates for PS in LN scenario with handedness as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Handedness	Ambidextrous	87.15	10.538	66.13	108.17
	Left-handed	58.03	6.899	44.28	71.79
	Right-handed	69.00	2.300	64.41	73.58
Inc.	SS	73.41	5.382	62.68	84.14
	GP	77.13	5.150	66.85	87.40
	TS	63.65	4.670	54.34	72.96

continued . . .

Table A.114: . . . continued

Effect		M	SE	LB	UB
Ambidextrous	* SS	94.40	13.290	67.89	120.90
	* GP	87.65	12.717	62.29	113.01
	* TS	79.41	11.530	56.41	102.41
Left-handed	* SS	50.67	8.700	33.32	68.02
	* GP	72.46	8.325	55.86	89.07
	* TS	50.97	7.548	35.91	66.02
Right-handed	* SS	75.16	2.900	69.38	80.94
	* GP	71.26	2.775	65.73	76.80
	* TS	60.57	2.516	55.55	65.59

Table A.115: Pairwise comparisons of main effects of inceptor and hand, based on estimated marginal means for PS in LN scenario. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
Handedness	Ambidextrous	Left-handed	29.118	12.595	.071	-1.777	60.012
		Right-handed	18.153	10.786	.290	-8.303	44.610
	Left-handed	Ambidextrous	-29.118	12.595	.071	-60.012	1.777
		Right-handed	-10.964	7.272	.408	-28.801	6.873
	Right-handed	Ambidextrous	-18.153	10.786	.290	-44.610	8.303
		Left-handed	10.964	7.272	.408	-6.873	28.801
Inc.	SS	GP	-3.715	4.471	1.000	-14.681	7.250
		TS	9.760	5.184	.192	-2.955	22.475
	GP	SS	3.715	4.471	1.000	-7.250	14.681
		TS	13.475	4.595	.014	2.204	24.747
	TS	SS	-9.760	5.184	.192	-22.475	2.955
		GP	-13.475	4.595	.014	-24.747	-2.204

Table A.116: Pairwise comparisons based on estimated marginal means for PS in LN scenario with handedness as a grouping factor showing patterns of interaction of inceptor and hand. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	Ambidextrous	Left-handed	43.724	15.884	.023	4.762	82.685
		Right-handed	19.237	13.602	.485	-14.128	52.601
	Left-handed	Ambidextrous	-43.724	15.884	.023	-82.685	-4.762
		Right-handed	-24.487	9.171	.028	-46.982	-1.993
	Right-handed	Ambidextrous	-19.237	13.602	.485	-52.601	14.128
		Left-handed	24.487	9.171	.028	1.993	46.982
GP	Ambidextrous	Left-handed	15.186	15.200	.964	-22.097	52.469
		Right-handed	16.386	13.016	.637	-15.542	48.313
	Left-handed	Ambidextrous	-15.186	15.200	.964	-52.469	22.097
		Right-handed	1.200	8.776	1.000	-20.326	22.725
	Right-handed	Ambidextrous	-16.386	13.016	.637	-48.313	15.542
		Left-handed	-1.200	8.776	1.000	-22.725	20.326
TS	Ambidextrous	Left-handed	28.443	13.781	.128	-5.360	62.247
		Right-handed	18.838	11.802	.345	-10.110	47.786
	Left-handed	Ambidextrous	-28.443	13.781	.128	-62.247	5.360
		Right-handed	-9.605	7.957	.694	-29.122	9.911
	Right-handed	Ambidextrous	-18.838	11.802	.345	-47.786	10.110
		Left-handed	9.605	7.957	.694	-9.911	29.122

Table A.117: Pairwise comparisons based on estimated marginal means for PS in LN scenario with handedness as a grouping factor showing patterns of interaction of handedness and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
Ambidextrous	SS	GP	6.748	11.039	1.000	-20.328	33.824

continued . . .

Table A.117: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
Ambidextrous	SS	TS	14.986	12.799	.737	-16.408	46.381
	GP	SS	-6.748	11.039	1.000	-33.824	20.328
		TS	8.239	11.347	1.000	-19.593	36.070
	TS	SS	-14.986	12.799	.737	-46.381	16.408
		GP	-8.239	11.347	1.000	-36.070	19.593
Left-handed	SS	GP	-21.790	7.226	.011	-39.516	-4.065
		TS	-.294	8.379	1.000	-20.847	20.259
	GP	SS	21.790	7.226	.011	4.065	39.516
		TS	21.496	7.428	.015	3.276	39.717
	TS	SS	.294	8.379	1.000	-20.259	20.847
		GP	-21.496	7.428	.015	-39.717	-3.276
Right-handed	SS	GP	3.897	2.409	.331	-2.012	9.805
		TS	14.588	2.793	$< .001$	7.737	21.439
	GP	SS	-3.897	2.409	.331	-9.805	2.012
		TS	10.691	2.476	$< .001$	4.618	16.765
	TS	SS	-14.588	2.793	$< .001$	-21.439	-7.737
		GP	-10.691	2.476	$< .001$	-16.765	-4.618

Table A.118: Post-hoc pairwise comparisons based on observed means with different methods for PS in LN scenario with handedness grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 333.145$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	Ambidextrous	Left-handed	29.118	12.595	.061	-1.042	59.278
		Right-handed	18.153	10.786	.219	-7.674	43.981
	Left-handed	Ambidextrous	-29.118	12.595	.061	-59.278	1.042
		Right-handed	-10.964	7.272	.294	-28.377	6.449
	Right-handed	Ambidextrous	-18.153	10.786	.219	-43.981	7.674
		Left-handed	10.964	7.272	.294	-6.449	28.377

continued . . .

Table A.118: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Bonferroni	Ambidextrous	Left-handed	29.118	12.595	.071	-1.777	60.012
		Right-handed	18.153	10.786	.290	-8.303	44.610
	Left-handed	Ambidextrous	-29.118	12.595	.071	-60.012	1.777
		Right-handed	-10.964	7.272	.408	-28.801	6.873
	Right-handed	Ambidextrous	-18.153	10.786	.290	-44.610	8.303
		Left-handed	10.964	7.272	.408	-6.873	28.801

ANOVA PS LN with GP

Table A.119: Estimates for PS in LN scenario with GP as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
GP usage	no	66.07	3.625	58.84	73.30
	hardly ever	67.60	5.034	57.55	77.64
	sometimes	69.32	4.212	60.92	77.72
	a lot	74.82	5.438	63.98	85.67
Inc.	SS	73.71	3.004	67.72	79.70
	GP	73.00	2.733	67.55	78.45
	TS	61.65	2.493	56.68	66.62
no	* SS	72.80	4.702	63.41	82.18
	* GP	69.15	4.278	60.62	77.68
	* TS	56.27	3.902	48.48	64.05
hardly ever	* SS	67.24	6.530	54.21	80.26
	* GP	74.49	5.941	62.64	86.34
	* TS	61.07	5.419	50.26	71.88
sometimes	* SS	76.05	5.464	65.15	86.95
	* GP	71.19	4.970	61.27	81.10

continued . . .

Table A.119: ...continued

Effect		M	SE	LB	UB
sometimes	* TS	60.73	4.534	51.68	69.77
a lot	* SS	78.77	7.053	64.69	92.84
	* GP	77.18	6.417	64.38	89.98
	* TS	68.53	5.853	56.85	80.21

Table A.120: Pairwise comparisons of main effects of inceptor and GP, based on estimated marginal means for PS in LN scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
GP usage	no	hardly ever	-1.527	6.204	1.000	-18.379	15.325
		sometimes	-3.252	5.557	1.000	-18.347	11.844
		a lot	-8.754	6.535	1.000	-26.507	8.998
	hardly ever	no	1.527	6.204	1.000	-15.325	18.379
		sometimes	-1.725	6.564	1.000	-19.555	16.106
		a lot	-7.227	7.410	1.000	-27.357	12.902
	sometimes	no	3.252	5.557	1.000	-11.844	18.347
		hardly ever	1.725	6.564	1.000	-16.106	19.555
		a lot	-5.503	6.878	1.000	-24.187	13.181
	a lot	no	8.754	6.535	1.000	-8.998	26.507
		hardly ever	7.227	7.410	1.000	-12.902	27.357
		sometimes	5.503	6.878	1.000	-13.181	24.187
Inc.	SS	GP	.711	2.494	1.000	-5.409	6.831
		TS	12.065	2.754	< .001	5.308	18.821
	GP	SS	-.711	2.494	1.000	-6.831	5.409
		TS	11.354	2.458	< .001	5.322	17.385
	TS	SS	-12.065	2.754	< .001	-18.821	-5.308
		GP	-11.354	2.458	< .001	-17.385	-5.322

Table A.121: Pairwise comparisons based on estimated marginal means for PS in LN scenario with GP as a grouping factor showing patterns of interaction of inceptor and GP. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	no	hardly ever	5.559	8.047	1.000	-16.301	27.418
		sometimes	-3.256	7.208	1.000	-22.837	16.326
		a lot	-5.971	8.477	1.000	-28.999	17.057
	hardly ever	no	-5.559	8.047	1.000	-27.418	16.301
		sometimes	-8.815	8.514	1.000	-31.943	14.314
		a lot	-11.529	9.612	1.000	-37.641	14.582
	sometimes	no	3.256	7.208	1.000	-16.326	22.837
		hardly ever	8.815	8.514	1.000	-14.314	31.943
		a lot	-2.715	8.922	1.000	-26.951	21.521
	a lot	no	5.971	8.477	1.000	-17.057	28.999
		hardly ever	11.529	9.612	1.000	-14.582	37.641
		sometimes	2.715	8.922	1.000	-21.521	26.951
GP	no	hardly ever	-5.341	7.321	1.000	-25.227	14.545
		sometimes	-2.039	6.558	1.000	-19.852	15.775
		a lot	-8.029	7.712	1.000	-28.978	12.920
	hardly ever	no	5.341	7.321	1.000	-14.545	25.227
		sometimes	3.302	7.746	1.000	-17.738	24.343
		a lot	-2.688	8.744	1.000	-26.442	21.066
	sometimes	no	2.039	6.558	1.000	-15.775	19.852
		hardly ever	-3.302	7.746	1.000	-24.343	17.738
		a lot	-5.990	8.116	1.000	-28.038	16.058
	a lot	no	8.029	7.712	1.000	-12.920	28.978
		hardly ever	2.688	8.744	1.000	-21.066	26.442
		sometimes	5.990	8.116	1.000	-16.058	28.038
TS	no	hardly ever	-4.799	6.678	1.000	-22.938	13.340

continued ...

Table A.121: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
TS	no	sometimes	-4.461	5.982	1.000	-20.709	11.788
		a lot	-12.264	7.034	.514	-31.372	6.845
	hardly ever	no	4.799	6.678	1.000	-13.340	22.938
		sometimes	.339	7.065	1.000	-18.854	19.531
		a lot	-7.464	7.976	1.000	-29.131	14.203
	sometimes	no	4.461	5.982	1.000	-11.788	20.709
		hardly ever	-.339	7.065	1.000	-19.531	18.854
		a lot	-7.803	7.404	1.000	-27.914	12.308
	a lot	no	12.264	7.034	.514	-6.845	31.372
		hardly ever	7.464	7.976	1.000	-14.203	29.131
		sometimes	7.803	7.404	1.000	-12.308	27.914

Table A.122: Pairwise comparisons based on estimated marginal means for PS in LN scenario with GP as a grouping factor showing patterns of interaction of GP and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
no	SS	GP	3.646	3.905	1.000	-5.935	13.228
		TS	16.529	4.311	.001	5.951	27.106
	GP	SS	-3.646	3.905	1.000	-13.228	5.935
		TS	12.882	3.848	.004	3.440	22.325
	TS	SS	-16.529	4.311	.001	-27.106	-5.951
		GP	-12.882	3.848	.004	-22.325	-3.440
hardly ever	SS	GP	-7.254	5.423	.556	-20.560	6.052
		TS	6.170	5.987	.919	-8.519	20.860
	GP	SS	7.254	5.423	.556	-6.052	20.560
		TS	13.424	5.344	.043	.311	26.537
	TS	SS	-6.170	5.987	.919	-20.860	8.519

continued . . .

Table A.122: ... continued

Subgroup	Inc.		MD	SE	p	LB	UB
hardly ever	TS	GP	-13.424	5.344	.043	-26.537	-.311
sometimes	SS	GP	4.863	4.537	.862	-6.269	15.996
		TS	15.324	5.009	.009	3.033	27.614
	GP	SS	-4.863	4.537	.862	-15.996	6.269
		TS	10.460	4.471	.067	-.510	21.431
	TS	SS	-15.324	5.009	.009	-27.614	-3.033
		GP	-10.460	4.471	.067	-21.431	.510
a lot	SS	GP	1.588	5.857	1.000	-12.784	15.960
		TS	10.236	6.466	.354	-5.631	26.102
	GP	SS	-1.588	5.857	1.000	-15.960	12.784
		TS	8.648	5.772	.416	-5.515	22.811
	TS	SS	-10.236	6.466	.354	-26.102	5.631
		GP	-8.648	5.772	.416	-22.811	5.515

Table A.123: Post-hoc pairwise comparisons based on observed means with different methods for PS in LN scenario with GP grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 354.810$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	no	hardly ever	-1.527	6.204	.995	-17.860	14.805
		sometimes	-3.252	5.557	.936	-17.882	11.379
		a lot	-8.754	6.535	.541	-25.960	8.451
	hardly ever	no	1.527	6.204	.995	-14.805	17.860
		sometimes	-1.725	6.564	.994	-19.006	15.557
		a lot	-7.227	7.410	.764	-26.737	12.282
	sometimes	no	3.252	5.557	.936	-11.379	17.882
		hardly ever	1.725	6.564	.994	-15.557	19.006
		a lot	-5.503	6.878	.854	-23.611	12.606

continued ...

Table A.123: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	a lot	no	8.754	6.535	.541	-8.451	25.960
		hardly ever	7.227	7.410	.764	-12.282	26.737
		sometimes	5.503	6.878	.854	-12.606	23.611
Bonferroni	no	hardly ever	-1.527	6.204	1.000	-18.379	15.325
		sometimes	-3.252	5.557	1.000	-18.347	11.844
		a lot	-8.754	6.535	1.000	-26.507	8.998
	hardly ever	no	1.527	6.204	1.000	-15.325	18.379
		sometimes	-1.725	6.564	1.000	-19.555	16.106
		a lot	-7.227	7.410	1.000	-27.357	12.902
	sometimes	no	3.252	5.557	1.000	-11.844	18.347
		hardly ever	1.725	6.564	1.000	-16.106	19.555
		a lot	-5.503	6.878	1.000	-24.187	13.181
	a lot	no	8.754	6.535	1.000	-8.998	26.507
		hardly ever	7.227	7.410	1.000	-12.902	27.357
		sometimes	5.503	6.878	1.000	-13.181	24.187

ANOVA PS LN with MG

Table A.124: Estimates for PS in LN scenario with MG as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
MG usage	no / h. ever	74.12	3.417	67.31	80.94
	used to	64.41	4.115	56.20	72.61
	yes	65.71	3.756	58.21	73.20
Inc.	SS	72.11	2.656	66.82	77.41
	GP	71.88	2.559	66.78	76.98

continued . . .

Table A.124: . . . continued

Effect		M	SE	LB	UB
Inc.	TS	60.24	2.431	55.39	65.09
no / h. ever	* SS	84.07	4.167	75.76	92.38
	* GP	76.52	4.013	68.52	84.53
	* TS	61.78	3.813	54.17	69.38
used to	* SS	59.88	5.017	49.87	69.89
	* GP	74.53	4.832	64.89	84.16
	* TS	58.81	4.592	49.65	67.97
yes	* SS	72.39	4.580	63.26	81.53
	* GP	64.59	4.411	55.79	73.39
	* TS	60.14	4.192	51.78	68.50

Table A.125: Pairwise comparisons of main effects of inceptor and MG, based on estimated marginal means for PS in LN scenario. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
MG usage	no / h. ever	used to	9.717	5.349	.221	-3.403	22.838
		yes	8.417	5.078	.306	-4.040	20.873
	used to	no / h. ever	-9.717	5.349	.221	-22.838	3.403
		yes	-1.301	5.572	1.000	-14.968	12.366
	yes	no / h. ever	-8.417	5.078	.306	-20.873	4.040
		used to	1.301	5.572	1.000	-12.366	14.968
Inc.	SS	GP	.235	2.127	1.000	-4.983	5.453
		TS	11.872	2.469	< .001	5.816	17.929
	GP	SS	-.235	2.127	1.000	-5.453	4.983
		TS	11.637	2.280	< .001	6.046	17.229
	TS	SS	-11.872	2.469	< .001	-17.929	-5.816
		GP	-11.637	2.280	< .001	-17.229	-6.046

Table A.126: Pairwise comparisons based on estimated marginal means for PS in LN scenario with MG as a grouping factor showing patterns of interaction of inceptor and MG. Inc. - inceptor; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	*p*	LB	UB
SS	no / h. ever	used to	24.188	6.522	.001	8.191	40.185
		yes	11.678	6.192	.190	-3.510	26.865
	used to	no / h. ever	-24.188	6.522	.001	-40.185	-8.191
		yes	-12.510	6.793	.209	-29.173	4.153
	yes	no / h. ever	-11.678	6.192	.190	-26.865	3.510
		used to	12.510	6.793	.209	-4.153	29.173
GP	no / h. ever	used to	1.996	6.281	1.000	-13.412	17.404
		yes	11.935	5.964	.148	-2.693	26.563
	used to	no / h. ever	-1.996	6.281	1.000	-17.404	13.412
		yes	9.939	6.543	.400	-6.110	25.989
	yes	no / h. ever	-11.935	5.964	.148	-26.563	2.693
		used to	-9.939	6.543	.400	-25.989	6.110
TS	no / h. ever	used to	2.969	5.969	1.000	-11.672	17.610
		yes	1.637	5.667	1.000	-12.263	15.537
	used to	no / h. ever	-2.969	5.969	1.000	-17.610	11.672
		yes	-1.332	6.218	1.000	-16.583	13.919
	yes	no / h. ever	-1.637	5.667	1.000	-15.537	12.263
		used to	1.332	6.218	1.000	-13.919	16.583

Table A.127: Pairwise comparisons based on estimated marginal means for PS in LN scenario with MG as a grouping factor showing patterns of interaction of MG and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	*p*	LB	UB
no / h. ever	SS	GP	7.546	3.337	.081	-.639	15.731

continued …

Table A.127: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
no / h. ever	SS	TS	22.292	3.873	$< .001$	12.792	31.792
	GP	SS	-7.546	3.337	.081	-15.731	.639
		TS	14.746	3.576	$< .001$	5.975	23.516
	TS	SS	-22.292	3.873	$< .001$	-31.792	-12.792
		GP	-14.746	3.576	$< .001$	-23.516	-5.975
used to	SS	GP	-14.646	4.018	.002	-24.501	-4.790
		TS	1.073	4.664	1.000	-10.366	12.512
	GP	SS	14.646	4.018	.002	4.790	24.501
		TS	15.719	4.306	.002	5.158	26.280
	TS	SS	-1.073	4.664	1.000	-12.512	10.366
		GP	-15.719	4.306	.002	-26.280	-5.158
yes	SS	GP	7.804	3.668	.111	-1.193	16.801
		TS	12.251	4.257	.016	1.809	22.694
	GP	SS	-7.804	3.668	.111	-16.801	1.193
		TS	4.447	3.930	.785	-5.194	14.088
	TS	SS	-12.251	4.257	.016	-22.694	-1.809
		GP	-4.447	3.930	.785	-14.088	5.194

Table A.128: Post-hoc pairwise comparisons based on observed means with different methods for PS in LN scenario with MG grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 338.667$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	no / h. ever	used to	9.717	5.349	.172	-3.091	22.526
		yes	8.417	5.078	.229	-3.744	20.577
	used to	no / h. ever	-9.717	5.349	.172	-22.526	3.091
		yes	-1.301	5.572	.970	-14.643	12.041
	yes	no / h. ever	-8.417	5.078	.229	-20.577	3.744
		used to	1.301	5.572	.970	-12.041	14.643

continued . . .

Table A.128: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Bonferroni	no / h. ever	used to	9.717	5.349	.221	-3.403	22.838
		yes	8.417	5.078	.306	-4.040	20.873
	used to	no / h. ever	-9.717	5.349	.221	-22.838	3.403
		yes	-1.301	5.572	1.000	-14.968	12.366
	yes	no / h. ever	-8.417	5.078	.306	-20.873	4.040
		used to	1.301	5.572	1.000	-12.366	14.968

ANOVA PS LN with inceptor order

Table A.129: Estimates for PS in LN scenario with inceptor order as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Inc. order	123	60.36	5.158	50.06	70.65
	132	68.50	5.369	57.79	79.22
	213	67.47	5.158	57.18	77.77
	231	69.73	5.607	58.54	80.93
	312	78.37	5.607	67.18	89.56
	321	69.36	5.158	59.06	79.65
Inc.	SS	74.06	2.729	68.61	79.51
	GP	72.30	2.512	67.28	77.31
	TS	60.54	2.441	55.67	65.42
123	* SS	60.39	6.449	47.52	73.26
	* GP	63.42	5.936	51.57	75.26
	* TS	57.27	5.768	45.76	68.78
132	* SS	64.73	6.713	51.33	78.13
	* GP	79.96	6.178	67.63	92.29
	* TS	60.82	6.003	48.84	72.81

continued . . .

Table A.129: . . . continued

Effect		M	SE	LB	UB
213	* SS	74.11	6.449	61.23	86.98
	* GP	67.79	5.936	55.94	79.63
	* TS	60.52	5.768	49.01	72.03
231	* SS	85.30	7.011	71.30	99.29
	* GP	65.50	6.453	52.63	78.38
	* TS	58.40	6.270	45.88	70.91
312	* SS	84.09	7.011	70.09	98.08
	* GP	84.14	6.453	71.26	97.02
	* TS	66.88	6.270	54.37	79.40
321	* SS	75.74	6.449	62.86	88.61
	* GP	72.96	5.936	61.12	84.81
	* TS	59.37	5.768	47.85	70.88

Table A.130: Pairwise comparisons of main effects of inceptor and inceptor order, based on estimated marginal means for PS in LN scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Inc. order	123	132	-8.145	7.445	1.000	-30.808	14.518
		213	-7.114	7.294	1.000	-29.319	15.091
		231	-9.375	7.619	1.000	-32.567	13.817
		312	-18.012	7.619	.315	-41.205	5.180
		321	-8.997	7.294	1.000	-31.202	13.208
	132	123	8.145	7.445	1.000	-14.518	30.808
		213	1.032	7.445	1.000	-21.631	23.695
		231	-1.230	7.763	1.000	-24.861	22.402
		312	-9.867	7.763	1.000	-33.498	13.764
		321	-.852	7.445	1.000	-23.515	21.811
	213	123	7.114	7.294	1.000	-15.091	29.319

continued . . .

Table A.130: . . . continued

Factor			MD	SE	p	LB	UB
Inc. order	213	132	-1.032	7.445	1.000	-23.695	21.631
		231	-2.261	7.619	1.000	-25.454	20.931
		312	-10.899	7.619	1.000	-34.091	12.294
		321	-1.884	7.294	1.000	-24.089	20.321
	231	123	9.375	7.619	1.000	-13.817	32.567
		132	1.230	7.763	1.000	-22.402	24.861
		213	2.261	7.619	1.000	-20.931	25.454
		312	-8.637	7.930	1.000	-32.777	15.502
		321	.378	7.619	1.000	-22.815	23.570
	312	123	18.012	7.619	.315	-5.180	41.205
		132	9.867	7.763	1.000	-13.764	33.498
		213	10.899	7.619	1.000	-12.294	34.091
		231	8.637	7.930	1.000	-15.502	32.777
		321	9.015	7.619	1.000	-14.178	32.207
	321	123	8.997	7.294	1.000	-13.208	31.202
		132	.852	7.445	1.000	-21.811	23.515
		213	1.884	7.294	1.000	-20.321	24.089
		231	-.378	7.619	1.000	-23.570	22.815
		312	-9.015	7.619	1.000	-32.207	14.178
Inc.	SS	GP	1.762	2.135	1.000	-3.481	7.005
		TS	13.514	2.528	$< .001$	7.307	19.721
	GP	SS	-1.762	2.135	1.000	-7.005	3.481
		TS	11.751	2.306	$< .001$	6.088	17.415
	TS	SS	-13.514	2.528	$< .001$	-19.721	-7.307
		GP	-11.751	2.306	$< .001$	-17.415	-6.088

Table A.131: Pairwise comparisons based on estimated marginal means for PS in LN scenario with inceptor order as a grouping factor showing patterns of interaction of inceptor and inceptor order. Inc. - inceptor; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	*p*	LB	UB
SS	123	132	-4.339	9.309	1.000	-32.675	23.997
		213	-13.718	9.121	1.000	-41.482	14.045
		231	-24.910	9.526	.165	-53.908	4.089
		312	-23.700	9.526	.230	-52.698	5.299
		321	-15.348	9.121	1.000	-43.112	12.416
	132	123	4.339	9.309	1.000	-23.997	32.675
		213	-9.379	9.309	1.000	-37.716	18.957
		231	-20.571	9.706	.567	-50.118	8.976
		312	-19.361	9.706	.752	-48.908	10.186
		321	-11.009	9.309	1.000	-39.345	17.328
	213	123	13.718	9.121	1.000	-14.045	41.482
		132	9.379	9.309	1.000	-18.957	37.716
		231	-11.191	9.526	1.000	-40.190	17.807
		312	-9.981	9.526	1.000	-38.980	19.017
		321	-1.629	9.121	1.000	-29.393	26.135
	231	123	24.910	9.526	.165	-4.089	53.908
		132	20.571	9.706	.567	-8.976	50.118
		213	11.191	9.526	1.000	-17.807	40.190
		312	1.210	9.915	1.000	-28.972	31.392
		321	9.562	9.526	1.000	-19.436	38.560
	312	123	23.700	9.526	.230	-5.299	52.698
		132	19.361	9.706	.752	-10.186	48.908
		213	9.981	9.526	1.000	-19.017	38.980
		231	-1.210	9.915	1.000	-31.392	28.972
		321	8.352	9.526	1.000	-20.646	37.350
	321	123	15.348	9.121	1.000	-12.416	43.112

continued . . .

Table A.131: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	321	132	11.009	9.309	1.000	-17.328	39.345
		213	1.629	9.121	1.000	-26.135	29.393
		231	-9.562	9.526	1.000	-38.560	19.436
		312	-8.352	9.526	1.000	-37.350	20.646
GP	123	132	-16.545	8.567	.865	-42.625	9.534
		213	-4.372	8.394	1.000	-29.924	21.181
		231	-2.089	8.767	1.000	-28.778	24.600
		312	-20.723	8.767	.315	-47.412	5.966
		321	-9.548	8.394	1.000	-35.100	16.005
	132	123	16.545	8.567	.865	-9.534	42.625
		213	12.173	8.567	1.000	-13.906	38.253
		231	14.456	8.933	1.000	-12.738	41.650
		312	-4.178	8.933	1.000	-31.371	23.016
		321	6.998	8.567	1.000	-19.082	33.077
	213	123	4.372	8.394	1.000	-21.181	29.924
		132	-12.173	8.567	1.000	-38.253	13.906
		231	2.283	8.767	1.000	-24.406	28.971
		312	-16.351	8.767	.998	-43.040	10.338
		321	-5.176	8.394	1.000	-30.728	20.377
	231	123	2.089	8.767	1.000	-24.600	28.778
		132	-14.456	8.933	1.000	-41.650	12.738
		213	-2.283	8.767	1.000	-28.971	24.406
		312	-18.634	9.125	.676	-46.412	9.145
		321	-7.458	8.767	1.000	-34.147	19.230
	312	123	20.723	8.767	.315	-5.966	47.412
		132	4.178	8.933	1.000	-23.016	31.371
		213	16.351	8.767	.998	-10.338	43.040
		231	18.634	9.125	.676	-9.145	46.412
		321	11.175	8.767	1.000	-15.513	37.864

continued . . .

Table A.131: … continued

Inc.	Grouping factor		MD	SE	p	LB	UB
GP	321	123	9.548	8.394	1.000	-16.005	35.100
		132	-6.998	8.567	1.000	-33.077	19.082
		213	5.176	8.394	1.000	-20.377	30.728
		231	7.458	8.767	1.000	-19.230	34.147
		312	-11.175	8.767	1.000	-37.864	15.513
TS	123	132	-3.552	8.325	1.000	-28.895	21.791
		213	-3.251	8.157	1.000	-28.082	21.580
		231	-1.126	8.520	1.000	-27.061	24.809
		312	-9.614	8.520	1.000	-35.550	16.321
		321	-2.097	8.157	1.000	-26.928	22.734
	132	123	3.552	8.325	1.000	-21.791	28.895
		213	.301	8.325	1.000	-25.042	25.644
		231	2.426	8.681	1.000	-24< .001	28.852
		312	-6.062	8.681	1.000	-32.488	20.363
		321	1.455	8.325	1.000	-23.888	26.798
	213	123	3.251	8.157	1.000	-21.580	28.082
		132	-.301	8.325	1.000	-25.644	25.042
		231	2.124	8.520	1.000	-23.811	28.060
		312	-6.364	8.520	1.000	-32.299	19.572
		321	1.154	8.157	1.000	-23.677	25.985
	231	123	1.126	8.520	1.000	-24.809	27.061
		132	-2.426	8.681	1.000	-28.852	24< .001
		213	-2.124	8.520	1.000	-28.060	23.811
		312	-8.488	8.868	1.000	-35.482	18.506
		321	-.971	8.520	1.000	-26.906	24.965
	312	123	9.614	8.520	1.000	-16.321	35.550
		132	6.062	8.681	1.000	-20.363	32.488
		213	6.364	8.520	1.000	-19.572	32.299
		231	8.488	8.868	1.000	-18.506	35.482
		321	7.517	8.520	1.000	-18.418	33.453

continued …

Table A.131: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
TS	321	123	2.097	8.157	1.000	-22.734	26.928
		132	-1.455	8.325	1.000	-26.798	23.888
		213	-1.154	8.157	1.000	-25.985	23.677
		231	.971	8.520	1.000	-24.965	26.906
		312	-7.517	8.520	1.000	-33.453	18.418

Table A.132: Pairwise comparisons based on estimated marginal means for PS in LN scenario with inceptor order as a grouping factor showing patterns of interaction of inceptor order and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
123	SS	GP	-3.027	5.045	1.000	-15.417	9.362
		TS	3.118	5.973	1.000	-11.549	17.785
	GP	SS	3.027	5.045	1.000	-9.362	15.417
		TS	6.145	5.450	.791	-7.237	19.528
	TS	SS	-3.118	5.973	1.000	-17.785	11.549
		GP	-6.145	5.450	.791	-19.528	7.237
132	SS	GP	-15.233	5.251	.015	-28.129	-2.338
		TS	3.905	6.217	1.000	-11.361	19.172
	GP	SS	15.233	5.251	.015	2.338	28.129
		TS	19.139	5.672	.004	5.210	33.068
	TS	SS	-3.905	6.217	1.000	-19.172	11.361
		GP	-19.139	5.672	.004	-33.068	-5.210
213	SS	GP	6.319	5.045	.644	-6.070	18.709
		TS	13.586	5.973	.078	-1.082	28.253
	GP	SS	-6.319	5.045	.644	-18.709	6.070
		TS	7.266	5.450	.561	-6.116	20.649
	TS	SS	-13.586	5.973	.078	-28.253	1.082

continued . . .

Table A.132: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
213	TS	GP	-7.266	5.450	.561	-20.649	6.116
231	SS	GP	19.793	5.485	.002	6.325	33.262
		TS	26.901	6.493	$<.001$	10.956	42.847
	GP	SS	-19.793	5.485	.002	-33.262	-6.325
		TS	7.108	5.925	.703	-7.440	21.656
	TS	SS	-26.901	6.493	$<.001$	-42.847	-10.956
		GP	-7.108	5.925	.703	-21.656	7.440
312	SS	GP	-.050	5.485	1.000	-13.519	13.418
		TS	17.203	6.493	.030	1.258	33.149
	GP	SS	.050	5.485	1.000	-13.418	13.519
		TS	17.254	5.925	.015	2.706	31.802
	TS	SS	-17.203	6.493	.030	-33.149	-1.258
		GP	-17.254	5.925	.015	-31.802	-2.706
321	SS	GP	2.773	5.045	1.000	-9.617	15.162
		TS	16.369	5.973	.024	1.701	31.036
	GP	SS	-2.773	5.045	1.000	-15.162	9.617
		TS	13.596	5.450	.045	.213	26.978
	TS	SS	-16.369	5.973	.024	-31.036	-1.701
		GP	-13.596	5.450	.045	-26.978	-.213

Table A.133: Post-hoc pairwise comparisons based on observed means with different methods for PS in LN scenario with inceptor order grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 345.863$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	123	132	-8.145	7.445	.882	-29.987	13.697
		213	-7.114	7.294	.924	-28.514	14.287
		231	-9.375	7.619	.820	-31.727	12.977
		312	-18.012	7.619	.184	-40.364	4.340

continued . . .

Table A.133: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
	123	321	-8.997	7.294	.819	-30.398	12.403
	132	123	8.145	7.445	.882	-13.697	29.987
		213	1.032	7.445	1.000	-20.810	22.874
		231	-1.230	7.763	1.000	-24.005	21.545
		312	-9.867	7.763	.800	-32.642	12.908
		321	-.852	7.445	1.000	-22.694	20.990
	213	123	7.114	7.294	.924	-14.287	28.514
		132	-1.032	7.445	1.000	-22.874	20.810
		231	-2.261	7.619	1.000	-24.614	20.091
		312	-10.899	7.619	.709	-33.251	11.453
		321	-1.884	7.294	1.000	-23.284	19.517
Tukey HSD	231	123	9.375	7.619	.820	-12.977	31.727
		132	1.230	7.763	1.000	-21.545	24.005
		213	2.261	7.619	1.000	-20.091	24.614
		312	-8.637	7.930	.884	-31.902	14.628
		321	.378	7.619	1.000	-21.975	22.730
	312	123	18.012	7.619	.184	-4.340	40.364
		132	9.867	7.763	.800	-12.908	32.642
		213	10.899	7.619	.709	-11.453	33.251
		231	8.637	7.930	.884	-14.628	31.902
		321	9.015	7.619	.843	-13.337	31.367
	321	123	8.997	7.294	.819	-12.403	30.398
		132	.852	7.445	1.000	-20.990	22.694
		213	1.884	7.294	1.000	-19.517	23.284
		231	-.378	7.619	1.000	-22.730	21.975
		312	-9.015	7.619	.843	-31.367	13.337
Bonferroni	123	132	-8.145	7.445	1.000	-30.808	14.518
		213	-7.114	7.294	1.000	-29.319	15.091
		231	-9.375	7.619	1.000	-32.567	13.817

continued . . .

Table A.133: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
	123	312	-18.012	7.619	.315	-41.205	5.180
		321	-8.997	7.294	1.000	-31.202	13.208
	132	123	8.145	7.445	1.000	-14.518	30.808
		213	1.032	7.445	1.000	-21.631	23.695
		231	-1.230	7.763	1.000	-24.861	22.402
		312	-9.867	7.763	1.000	-33.498	13.764
		321	-.852	7.445	1.000	-23.515	21.811
	213	123	7.114	7.294	1.000	-15.091	29.319
		132	-1.032	7.445	1.000	-23.695	21.631
		231	-2.261	7.619	1.000	-25.454	20.931
		312	-10.899	7.619	1.000	-34.091	12.294
		321	-1.884	7.294	1.000	-24.089	20.321
Bonferroni	231	123	9.375	7.619	1.000	-13.817	32.567
		132	1.230	7.763	1.000	-22.402	24.861
		213	2.261	7.619	1.000	-20.931	25.454
		312	-8.637	7.930	1.000	-32.777	15.502
		321	.378	7.619	1.000	-22.815	23.570
	312	123	18.012	7.619	.315	-5.180	41.205
		132	9.867	7.763	1.000	-13.764	33.498
		213	10.899	7.619	1.000	-12.294	34.091
		231	8.637	7.930	1.000	-15.502	32.777
		321	9.015	7.619	1.000	-14.178	32.207
	321	123	8.997	7.294	1.000	-13.208	31.202
		132	.852	7.445	1.000	-21.811	23.515
		213	1.884	7.294	1.000	-20.321	24.089
		231	-.378	7.619	1.000	-23.570	22.815
		312	-9.015	7.619	1.000	-32.207	14.178

ANOVA PS LN with FE group

Table A.134: Estimates for PS in LN scenario with FE group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
FE gr.	A	76.30	4.032	68.26	84.34
	B	73.41	3.747	65.94	80.89
	C	60.93	3.107	54.73	67.13
Inc.	SS	75.87	2.664	70.56	81.18
	GP	73.91	2.452	69.02	78.80
	TS	60.87	2.452	55.98	65.76
A	* SS	86.61	5.099	76.44	96.77
	* GP	80.31	4.692	70.95	89.67
	* TS	61.99	4.692	52.63	71.35
B	* SS	78.46	4.738	69.01	87.91
	* GP	79.13	4.361	70.43	87.82
	* TS	62.66	4.360	53.96	71.35
C	* SS	62.54	3.929	54.71	70.38
	* GP	62.29	3.616	55.07	69.50
	* TS	57.96	3.615	50.75	65.17

Table A.135: Pairwise comparisons of main effects of inceptor and FE group, based on estimated marginal means for PS in LN scenario. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
FE gr.	A	B	2.887	5.505	1.000	-10.615	16.390
		C	15.371	5.091	.011	2.885	27.857
	B	A	-2.887	5.505	1.000	-16.390	10.615
		C	12.484	4.868	.037	.543	24.424
	C	A	-15.371	5.091	.011	-27.857	-2.885

continued . . .

Table A.135: . . . continued

Factor			MD	SE	p	LB	UB
FE gr.	C	B	-12.484	4.868	.037	-24.424	-.543
Inc.	SS	GP	1.963	2.450	1.000	-4.047	7.972
		TS	15.000	2.515	< .001	8.831	21.168
	GP	SS	-1.963	2.450	1.000	-7.972	4.047
		TS	13.037	2.253	< .001	7.510	18.565
	TS	SS	-15.000	2.515	< .001	-21.168	-8.831
		GP	-13.037	2.253	< .001	-18.565	-7.510

Table A.136: Pairwise comparisons based on estimated marginal means for PS in LN scenario with FE group as a grouping factor showing patterns of interaction of inceptor and FE group. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	A	B	8.148	6.960	.737	-8.925	25.221
		C	24.062	6.437	.001	8.273	39.850
	B	A	-8.148	6.960	.737	-25.221	8.925
		C	15.914	6.155	.035	.816	31.011
	C	A	-24.062	6.437	.001	-39.850	-8.273
		B	-15.914	6.155	.035	-31.011	-.816
GP	A	B	1.183	6.406	1.000	-14.530	16.895
		C	18.022	5.924	.010	3.492	32.553
	B	A	-1.183	6.406	1.000	-16.895	14.530
		C	16.839	5.665	.012	2.945	30.734
	C	A	-18.022	5.924	.010	-32.553	-3.492
		B	-16.839	5.665	.012	-30.734	-2.945
TS	A	B	-.669	6.405	1.000	-16.380	15.041
		C	4.029	5.923	1.000	-10.500	18.557
	B	A	.669	6.405	1.000	-15.041	16.380
		C	4.698	5.664	1.000	-9.195	18.591

continued . . .

Table A.136: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
TS	C	A	-4.029	5.923	1.000	-18.557	10.500
		B	-4.698	5.664	1.000	-18.591	9.195

Table A.137: Pairwise comparisons based on estimated marginal means for PS in LN scenario with FE group as a grouping factor showing patterns of interaction of FE group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
Female	SS	GP	-10.519	4.019	.033	-20.378	-.661
		TS	14.963	4.583	.005	3.723	26.204
	GP	SS	10.519	4.019	.033	.661	20.378
		TS	25.483	4.781	$< .001$	13.755	37.211
	TS	SS	-14.963	4.583	.005	-26.204	-3.723
		GP	-25.483	4.781	$< .001$	-37.211	-13.755
Male	SS	GP	-6.621	2.320	.017	-12.313	-.929
		TS	17.318	2.646	$< .001$	10.828	23.807
	GP	SS	6.621	2.320	.017	.929	12.313
		TS	23.938	2.760	$< .001$	17.167	30.709
	TS	SS	-17.318	2.646	$< .001$	-23.807	-10.828
		GP	-23.938	2.760	$< .001$	-30.709	-17.167
Pref. not to say	SS	GP	-7.896	17.052	1.000	-49.722	33.931
		TS	31.643	19.443	.324	-16.048	79.333
	GP	SS	7.896	17.052	1.000	-33.931	49.722
		TS	39.538	20.285	.166	-10.218	89.295
	TS	SS	-31.643	19.443	.324	-79.333	16.048
		GP	-39.538	20.285	.166	-89.295	10.218

Table A.138: Post-hoc pairwise comparisons based on observed means with different methods for PS in LN scenario with FE group grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 308.929$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	A	B	2.887	5.505	.860	-10.294	16.069
		C	15.371	5.091	.010	3.181	27.561
	B	A	-2.887	5.505	.860	-16.069	10.294
		C	12.484	4.868	.033	.827	24.140
	C	A	-15.371	5.091	.010	-27.561	-3.181
		B	-12.484	4.868	.033	-24.140	-.827
Bonferroni	A	B	2.887	5.505	1.000	-10.615	16.390
		C	15.371	5.091	.011	2.885	27.857
	B	A	-2.887	5.505	1.000	-16.390	10.615
		C	12.484	4.868	.037	.543	24.424
	C	A	-15.371	5.091	.011	-27.857	-2.885
		B	-12.484	4.868	.037	-24.424	-.543

ANOVA PS LD with gender

Table A.139: Descriptive statistics for PS in LD scenario with gender as a grouping factor. M - mean; SD - standard deviation; N - number of samples.

Subgroup	SS		GP		TS		N
	M	SD	M	SD	M	SD	
Female	49.38	28.56	54.43	33.56	47.68	27.28	18
Male	70.50	17.43	72.52	19.30	55.06	19.97	54
Pref. not to say	99.99		69.18		67.28		1
Total	65.69	22.73	68.01	24.52	53.41	21.96	73

Table A.140: Estimates for PS in LD scenario with gender as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
Gender	Female	50.49	4.266	41.99	59.00
	Male	66.03	2.463	61.11	70.94
	Pref. not to say	78.82	18.101	42.72	114.92
Inc.	SS	73.29	7.149	59.03	87.55
	GP	65.38	8.142	49.14	81.62
	TS	56.67	7.589	41.54	71.81
Female	* SS	49.38	4.878	39.65	59.11
	* GP	54.43	5.555	43.35	65.51
	* TS	47.68	5.178	37.35	58.00
Male	* SS	70.50	2.816	64.88	76.11
	* GP	72.52	3.207	66.12	78.92
	* TS	55.06	2.990	49.10	61.03
Pref. not to say	* SS	99.99	20.694	58.72	141.26
	* GP	69.18	23.568	22.18	116.19
	* TS	67.28	21.969	23.46	111.10

Table A.141: Pairwise comparisons of main effects of inceptor and gender, based on estimated marginal means for PS in LD scenario. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	p	LB	UB
Gender	Female	Male	-15.532	4.926	.007	-27.615	-3.448
		Pref. not to say	-28.324	18.597	.397	-73.939	17.291
	Male	Female	15.532	4.926	.007	3.448	27.615
		Pref. not to say	-12.792	18.267	1.000	-57.600	32.016
	Pref. not to say	Female	28.324	18.597	.397	-17.291	73.939

continued ...

Table A.141: . . . continued

Factor			MD	SE	*p*	LB	UB
Gender	Pref. not to say	Male	12.792	18.267	1.000	-32.016	57.600
Inc.	SS	GP	7.911	7.054	.798	-9.392	25.213
		TS	16.615	7.095	.066	-.788	34.019
	GP	SS	-7.911	7.054	.798	-25.213	9.392
		TS	8.705	8.543	.935	-12.251	29.660
	TS	SS	-16.615	7.095	.066	-34.019	.788
		GP	-8.705	8.543	.935	-29.660	12.251

Table A.142: Pairwise comparisons based on estimated marginal means for PS in LD scenario with gender as a grouping factor showing patterns of interaction of inceptor and gender. Inc. - inceptor; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	*p*	LB	UB
SS	Female	Male	-21.117	5.632	.001	-34.932	-7.302
		Pref. not to say	-50.611	21.261	.060	-102.763	1.540
	Male	Female	21.117	5.632	.001	7.302	34.932
		Pref. not to say	-29.494	20.885	.487	-80.722	21.735
	Pref. not to say	Female	50.611	21.261	.060	-1.540	102.763
		Male	29.494	20.885	.487	-21.735	80.722
GP	Female	Male	-18.090	6.414	.019	-33.823	-2.356
		Pref. not to say	-14.756	24.214	1.000	-74.149	44.638
	Male	Female	18.090	6.414	.019	2.356	33.823
		Pref. not to say	3.334	23.785	1.000	-55.009	61.676
	Pref. not to say	Female	14.756	24.214	1.000	-44.638	74.149
		Male	-3.334	23.785	1.000	-61.676	55.009
TS	Female	Male	-7.388	5.979	.662	-22.054	7.278
		Pref. not to say	-19.604	22.571	1.000	-74.968	35.759
	Male	Female	7.388	5.979	.662	-7.278	22.054
		Pref. not to say	-12.216	22.171	1.000	-66.600	42.167

continued . . .

Table A.142: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
TS	Pref. not to say	Female	19.604	22.571	1.000	-35.759	74.968
		Male	12.216	22.171	1.000	-42.167	66.600

Table A.143: Pairwise comparisons based on estimated marginal means for PS in LD scenario with gender as a grouping factor showing patterns of interaction of gender and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
Female	SS	GP	-5.050	4.813	.893	-16.855	6.755
		TS	1.703	4.841	1.000	-10.171	13.578
	GP	SS	5.050	4.813	.893	-6.755	16.855
		TS	6.754	5.829	.752	-7.544	21.051
	TS	SS	-1.703	4.841	1.000	-13.578	10.171
		GP	-6.754	5.829	.752	-21.051	7.544
Male	SS	GP	-2.023	2.779	1.000	-8.838	4.793
		TS	15.433	2.795	$< .001$	8.577	22.288
	GP	SS	2.023	2.779	1.000	-4.793	8.838
		TS	17.455	3.365	$< .001$	9.201	25.710
	TS	SS	-15.433	2.795	$< .001$	-22.288	-8.577
		GP	-17.455	3.365	$< .001$	-25.710	-9.201
Pref. not to say	SS	GP	30.805	20.419	.408	-19.280	80.891
		TS	32.710	20.538	.347	-17.668	83.088
	GP	SS	-30.805	20.419	.408	-80.891	19.280
		TS	1.905	24.730	1.000	-58.755	62.565
	TS	SS	-32.710	20.538	.347	-83.088	17.668
		GP	-1.905	24.730	1.000	-62.565	58.755

ANOVA PS LD with MG

Table A.144: Estimates for PS in LD scenario with MG as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
MG usage	no / h. ever	68.02	3.499	61.04	75.00
	used to	60.13	4.213	51.73	68.54
	yes	57.41	3.846	49.74	65.08
Inc.	SS	64.51	2.502	59.52	69.50
	GP	68.11	2.915	62.30	73.93
	TS	52.94	2.601	47.75	58.13
no / h. ever	* SS	76.79	3.925	68.96	84.61
	* GP	69.47	4.572	60.35	78.59
	* TS	57.79	4.080	49.66	65.93
used to	* SS	58.33	4.726	48.91	67.76
	* GP	71.64	5.505	60.66	82.62
	* TS	50.43	4.913	40.63	60.23
yes	* SS	58.42	4.314	49.82	67.03
	* GP	63.23	5.025	53.21	73.25
	* TS	50.59	4.485	41.65	59.54

Table A.145: Pairwise comparisons of main effects of inceptor and MG, based on estimated marginal means for PS in LD scenario. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
MG usage	no / h. ever	used to	7.883	5.477	.464	-5.551	21.316
		yes	10.604	5.199	.136	-2.149	23.358
	used to	no / h. ever	-7.883	5.477	.464	-21.316	5.551
		yes	2.721	5.705	1.000	-11.271	16.714
	yes	no / h. ever	-10.604	5.199	.136	-23.358	2.149

continued . . .

Table A.145: . . . continued

Factor			MD	SE	p	LB	UB
MG usage	yes	used to	-2.721	5.705	1.000	-16.714	11.271
Inc.	SS	GP	-3.599	2.245	.340	-9.105	1.908
		TS	11.575	2.464	$< .001$	5.530	17.620
	GP	SS	3.599	2.245	.340	-1.908	9.105
		TS	15.173	2.945	$< .001$	7.950	22.396
	TS	SS	-11.575	2.464	$< .001$	-17.620	-5.530
		GP	-15.173	2.945	$< .001$	-22.396	-7.950

Table A.146: Pairwise comparisons based on estimated marginal means for PS in LD scenario with MG as a grouping factor showing patterns of interaction of inceptor and MG. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	p	LB	UB
SS	no / h. ever	used to	18.452	6.143	.011	3.384	33.521
		yes	18.366	5.832	.007	4.060	32.672
	used to	no / h. ever	-18.452	6.143	.011	-33.521	-3.384
		yes	-.086	6.399	1.000	-15.782	15.610
	yes	no / h. ever	-18.366	5.832	.007	-32.672	-4.060
		used to	.086	6.399	1.000	-15.610	15.782
GP	no / h. ever	used to	-2.165	7.156	1.000	-19.717	15.388
		yes	6.243	6.794	1.000	-10.421	22.907
	used to	no / h. ever	2.165	7.156	1.000	-15.388	19.717
		yes	8.408	7.454	.790	-9.875	26.691
	yes	no / h. ever	-6.243	6.794	1.000	-22.907	10.421
		used to	-8.408	7.454	.790	-26.691	9.875
TS	no / h. ever	used to	7.361	6.386	.759	-8.304	23.025
		yes	7.203	6.063	.716	-7.668	22.075
	used to	no / h. ever	-7.361	6.386	.759	-23.025	8.304
		yes	-.157	6.652	1.000	-16.474	16.159

continued . . .

Table A.146: . . . continued

Inc.	Grouping factor		MD	SE	p	LB	UB
TS	yes	no / h. ever	-7.203	6.063	.716	-22.075	7.668
		used to	.157	6.652	1.000	-16.159	16.474

Table A.147: Pairwise comparisons based on estimated marginal means for PS in LD scenario with MG as a grouping factor showing patterns of interaction of MG and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	p	LB	UB
no / h. ever	SS	GP	7.315	3.521	.124	-1.323	15.952
		TS	18.993	3.865	$<$ .001	9.511	28.474
	GP	SS	-7.315	3.521	.124	-15.952	1.323
		TS	11.678	4.619	.041	.349	23.008
	TS	SS	-18.993	3.865	$<$ .001	-28.474	-9.511
		GP	-11.678	4.619	.041	-23.008	-.349
used to	SS	GP	-13.302	4.240	.007	-23.703	-2.902
		TS	7.901	4.654	.282	-3.516	19.318
	GP	SS	13.302	4.240	.007	2.902	23.703
		TS	21.203	5.562	.001	7.561	34.846
	TS	SS	-7.901	4.654	.282	-19.318	3.516
		GP	-21.203	5.562	.001	-34.846	-7.561
yes	SS	GP	-4.808	3.871	.655	-14.302	4.686
		TS	7.830	4.249	.209	-2.592	18.252
	GP	SS	4.808	3.871	.655	-4.686	14.302
		TS	12.638	5.077	.046	.184	25.092
	TS	SS	-7.830	4.249	.209	-18.252	2.592
		GP	-12.638	5.077	.046	-25.092	-.184

Table A.148: Post-hoc pairwise comparisons based on observed means with different methods for PS in LD scenario with MG grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 355.009$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	no / h. ever	used to	7.883	5.477	.327	-5.231	20.996
		yes	10.604	5.199	.110	-1.846	23.054
	used to	no / h. ever	-7.883	5.477	.327	-20.996	5.231
		yes	2.721	5.705	.882	-10.939	16.381
	yes	no / h. ever	-10.604	5.199	.110	-23.054	1.846
		used to	-2.721	5.705	.882	-16.381	10.939
Bonferroni	no / h. ever	used to	7.883	5.477	.464	-5.551	21.316
		yes	10.604	5.199	.136	-2.149	23.358
	used to	no / h. ever	-7.883	5.477	.464	-21.316	5.551
		yes	2.721	5.705	1.000	-11.271	16.714
	yes	no / h. ever	-10.604	5.199	.136	-23.358	2.149
		used to	-2.721	5.705	1.000	-16.714	11.271

ANOVA PS LD with FE group

Table A.149: Estimates for PS in LD scenario with FE group as a grouping factor. M - mean; SE - standard error; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval. Asterisks mean an interaction between the independent factor and the inceptor.

Effect		M	SE	LB	UB
FE gr.	A	73.36	3.926	65.53	81.19
	B	67.29	3.648	60.01	74.57
	C	52.46	3.025	46.43	58.50
Inc.	SS	68.35	2.339	63.69	73.02
	GP	70.11	2.720	64.69	75.53
	TS	54.65	2.572	49.52	59.78

continued ...

Table A.149: . . . continued

Effect		M	SE	LB	UB
A	* SS	81.32	4.476	72.39	90.25
	* GP	76.50	5.205	66.12	86.88
	* TS	62.27	4.923	52.45	72.09
B	* SS	70.92	4.160	62.63	79.22
	* GP	77.11	4.837	67.47	86.76
	* TS	53.83	4.575	44.71	62.96
C	* SS	52.82	3.449	45.94	59.70
	* GP	56.72	4.010	48.72	64.72
	* TS	47.86	3.793	40.29	55.42

Table A.150: Pairwise comparisons of main effects of inceptor and FE group, based on estimated marginal means for PS in LD scenario. MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Factor			MD	SE	*p*	LB	UB
FE gr.	A	B	6.074	5.359	.783	-7.072	19.219
		C	20.898	4.956	< .001	8.741	33.054
	B	A	-6.074	5.359	.783	-19.219	7.072
		C	14.824	4.739	.008	3.200	26.449
	C	A	-20.898	4.956	< .001	-33.054	-8.741
		B	-14.824	4.739	.008	-26.449	-3.200
Inc.	SS	GP	-1.756	2.442	1.000	-7.745	4.234
		TS	13.701	2.457	< .001	7.675	19.727
	GP	SS	1.756	2.442	1.000	-4.234	7.745
		TS	15.457	2.930	< .001	8.271	22.643
	TS	SS	-13.701	2.457	< .001	-19.727	-7.675
		GP	-15.457	2.930	< .001	-22.643	-8.271

Table A.151: Pairwise comparisons based on estimated marginal means for PS in LD scenario with FE group as a grouping factor showing patterns of interaction of inceptor and FE group. Inc. - inceptor; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Inc.	Grouping factor		MD	SE	*p*	LB	UB
SS	A	B	10.398	6.110	.280	-4.590	25.386
		C	28.501	5.651	< .001	14.641	42.362
	B	A	-10.398	6.110	.280	-25.386	4.590
		C	18.104	5.403	.004	4.850	31.358
	C	A	-28.501	5.651	< .001	-42.362	-14.641
		B	-18.104	5.403	.004	-31.358	-4.850
GP	A	B	-.615	7.105	1.000	-18.043	16.813
		C	19.782	6.571	.011	3.665	35.898
	B	A	.615	7.105	1.000	-16.813	18.043
		C	20.397	6.283	.005	4.985	35.808
	C	A	-19.782	6.571	.011	-35.898	-3.665
		B	-20.397	6.283	.005	-35.808	-4.985
TS	A	B	8.438	6.721	.640	-8.047	24.923
		C	14.411	6.215	.070	-.834	29.655
	B	A	-8.438	6.721	.640	-24.923	8.047
		C	5.973	5.943	.955	-8.605	20.551
	C	A	-14.411	6.215	.070	-29.655	.834
		B	-5.973	5.943	.955	-20.551	8.605

Table A.152: Pairwise comparisons based on estimated marginal means for PS in LD scenario with FE group as a grouping factor showing patterns of interaction of FE group and inceptor. Inc. - inceptor; MD - mean difference; SE - standard error; *p* - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference.

Subgroup	Inc.		MD	SE	*p*	LB	UB
A	SS	GP	4.822	4.673	.917	-6.640	16.284

continued . . .

Table A.152: . . . continued

Subgroup	Inc.		MD	SE	p	LB	UB
A	SS	TS	19.052	4.702	< .001	7.519	30.584
	GP	SS	-4.822	4.673	.917	-16.284	6.640
		TS	14.230	5.606	.040	.478	27.981
	TS	SS	-19.052	4.702	< .001	-30.584	-7.519
		GP	-14.230	5.606	.040	-27.981	-.478
B	SS	GP	-6.191	4.343	.475	-16.843	4.461
		TS	17.092	4.369	.001	6.374	27.809
	GP	SS	6.191	4.343	.475	-4.461	16.843
		TS	23.282	5.210	< .001	10.503	36.062
	TS	SS	-17.092	4.369	.001	-27.809	-6.374
		GP	-23.282	5.210	< .001	-36.062	-10.503
C	SS	GP	-3.898	3.601	.848	-12.730	4.934
		TS	4.961	3.623	.526	-3.926	13.847
	GP	SS	3.898	3.601	.848	-4.934	12.730
		TS	8.859	4.320	.132	-1.738	19.455
	TS	SS	-4.961	3.623	.526	-13.847	3.926
		GP	-8.859	4.320	.132	-19.455	1.738

Table A.153: Post-hoc pairwise comparisons based on observed means with different methods for PS in LD scenario with FE group grouping factor. MD - mean difference; SE - standard error; p - significance; LB - lower bound; UB - upper bound. LB and UB are in a 95% confidence interval for difference. The error term is Mean Square(Error) $MS_{Error} = 292.817$.

Method	Grouping factor		MD	SE	p	LB	UB
Tukey HSD	A	B	6.074	5.359	.497	-6.759	18.907
		C	20.898	4.956	< .001	9.030	32.765
	B	A	-6.074	5.359	.497	-18.907	6.759
		C	14.824	4.739	.007	3.476	26.173
	C	A	-20.898	4.956	< .001	-32.765	-9.030
		B	-14.824	4.739	.007	-26.173	-3.476

continued . . .

Table A.153: . . . continued

Method	Grouping factor		MD	SE	p	LB	UB
Bonferroni	A	B	6.074	5.359	.783	-7.072	19.219
		C	20.898	4.956	$<$ .001	8.741	33.054
	B	A	-6.074	5.359	.783	-19.219	7.072
		C	14.824	4.739	.008	3.200	26.449
	C	A	-20.898	4.956	$<$ .001	-33.054	-8.741
		B	-14.824	4.739	.008	-26.449	-3.200

Appendix B

Attached CD Content

The CD attached to this thesis contains the following:

1. This thesis in a PDF format;
2. Thesis summary;
3. Supplemental documents for the submission;
4. Cranfield University Research Ethics System (CURES) approval for the conducted trials.