UMTL

dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz *German Research Center for Artificial Intelligence*

SIC Saarland Informatics Campus

UNIVERSITÄT DES SAARLANDES

Seminar
Sweat and Survive - the VR Edition

Introduction to User Studies & Scientific Reports

May 27, 2025

Dr. Felix Kosmalla & Dr. André Zenner
Saarland University & DFKI

Goals of this Session

Understand the basics of planning and conducting a user study

Awareness of participant limitations in physical studies

Design, implement, and conduct a study using Unity-based VR apps

Get to know basic techniques for analyzing and reporting results

Part I: Understanding User Studies in HCI

Why User Studies Matter

"Works for me!"

But does it for your users?

Why User Studies Matter

"Works for me!"

But does it for your users?

Basic Study Concepts

Independent Variables

What you change

Dependent Variable

What you measure

- Performance
- Heart Rate
- UX
- …

Control Variables

What you keep constant

- Task
- Difficulty

Experimental Designs

Conditions

Danger Level

low

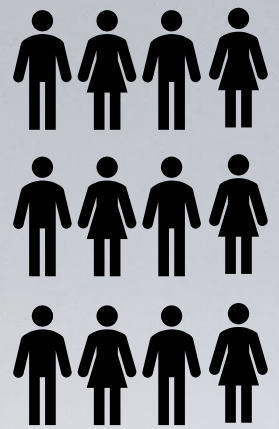
mid

high

?

Who tries which condition?

Participants



Experimental Designs

Within-Subject Design

- Each participant experiences all conditions

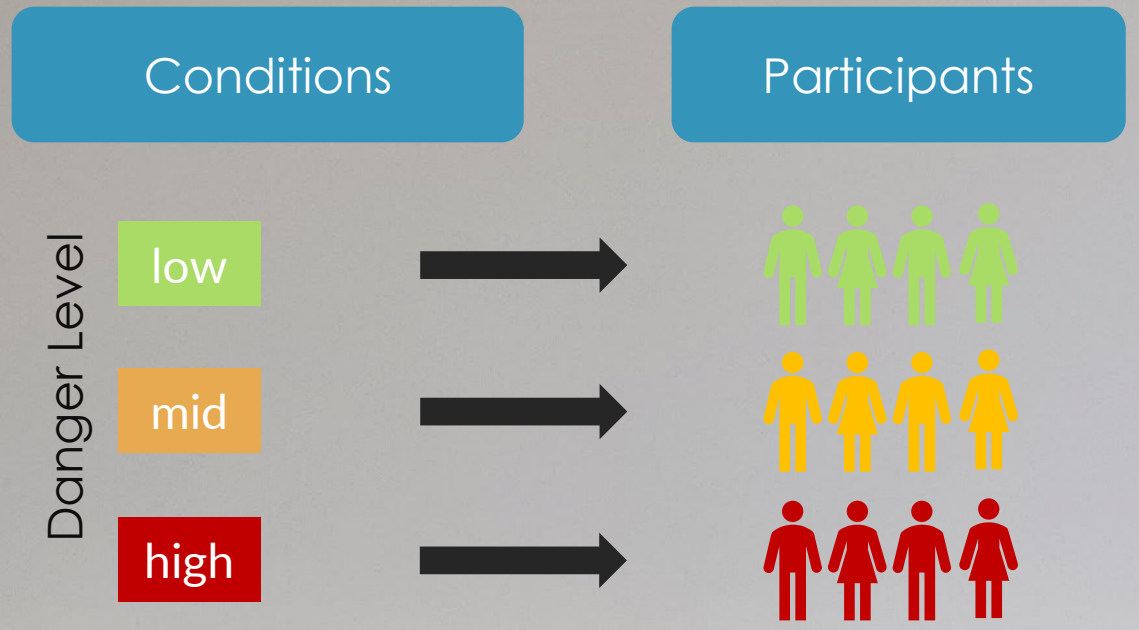
Pros

- Smaller N to find a (significant) difference
- Controls for individual differences
- Direct comparison

Cons

- Carryover effects
- Fatigue
- Learning effects
- Ordering effects *(tbc)*

Experimental Designs



Between-Subject Design

- Each participant experiences one condition

Pros	*Cons*
▶ No learning effects	▶ More Participants Needed
▶ No (multiplied) fatique	▶ Lower statistical power
▶ Less time commitment for participants	▶ Greater individual variability

Experimental Designs

Between-Subject Design

- Each participant experiences one condition

Pros

- No learning effects
- No (multiplied) fatique
- Less time commitment for participants

Cons

- More Participants Needed
- Lower statistical power
- Greater individual variability

Within-Subject Design

- Each participant experiences all conditions

Pros

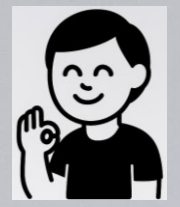
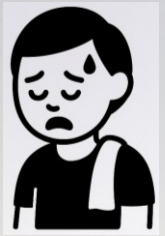
- Smaller N to find a (significant) difference
- Controls for individual differences
- Direct comparison

Cons

- Carryover effects
- Fatigue
- Learning effects
- Ordering effects

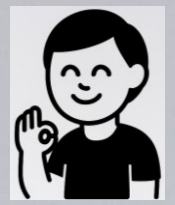
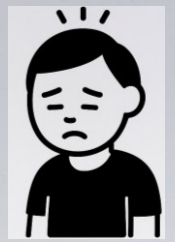
Counterbalancing

Ordering effects

Trial	Condition	Exhaustion
1st	low	
2nd	mid	
3rd	high	

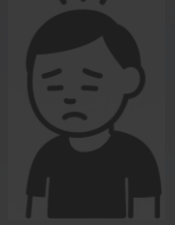
Counterbalancing

Ordering effects

Trial	Cond. P1	Cond. P2	Cond. P3	...	Exhaustion
1st	low	high	mid		
2nd	mid	low	high		
3rd	high	mid	low		

Counterbalancing

Ordering effects

Trial	Cond. P1	Cond. P2	Cond. P3	…	Exhaustion
1st	low	high	mid		
2nd	mid	low	high		
3rd	high	mid	low		

Best Case: **Fully Counterbalenced**

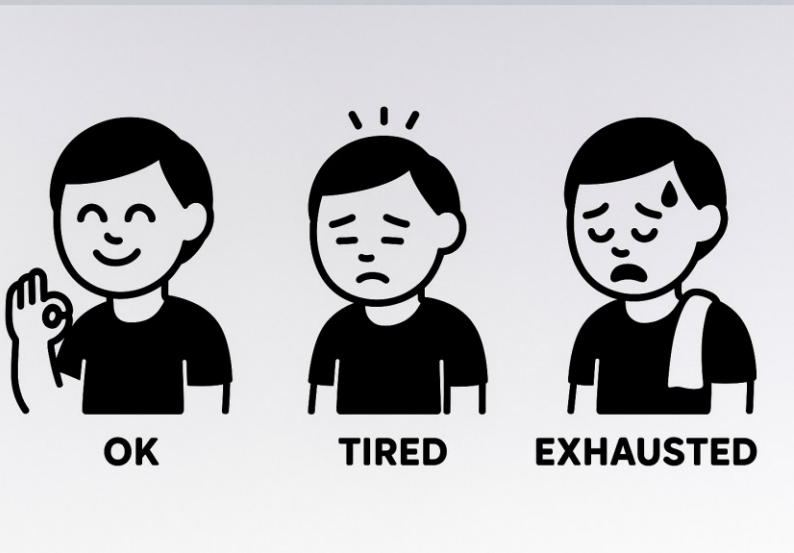
Multiples of 3!=3x2x1=6 Sequences

*3 conditions

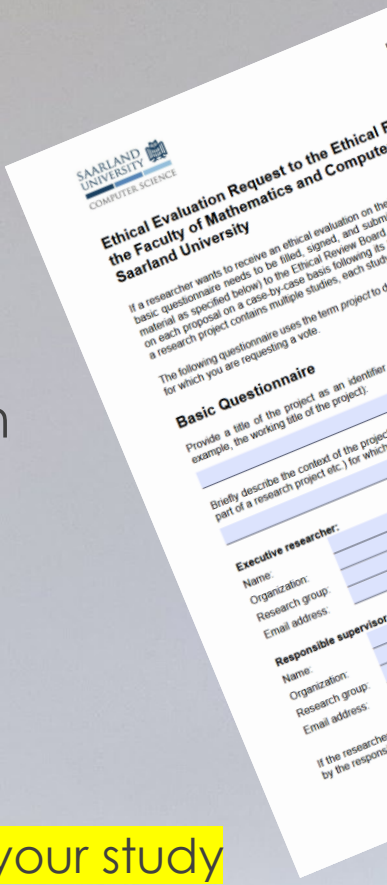
Part II: Working with Participants

Physical Strain in VR Sports Studies

- Sports applications = high physical load
- Be realistic: How many repetitions are reasonable?
- Give breaks, monitor well-being
- Have a plan for stopping if needed

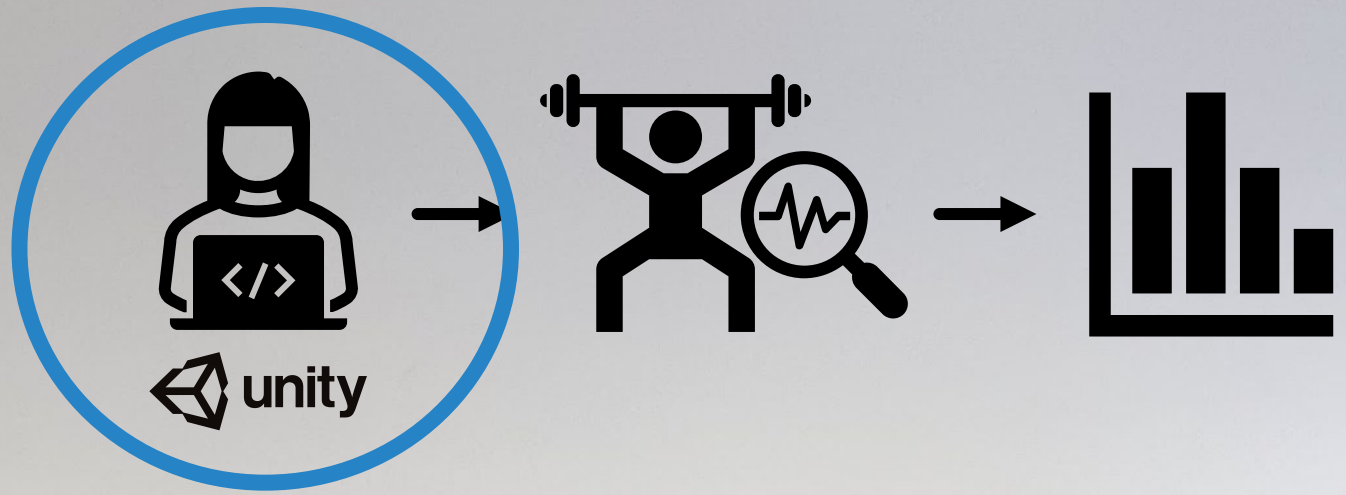


Ethics and Data Protection

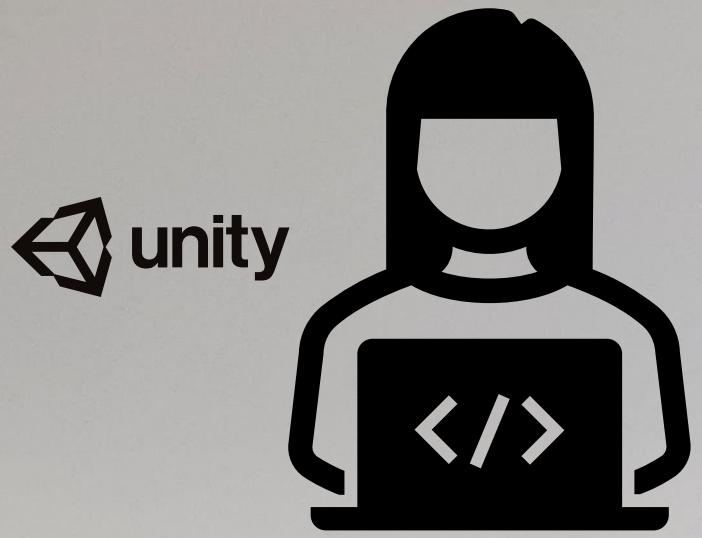


- Studies with human subjects need an approval of the Ethics Commission
  - in this seminar, we handle this for you
  - https://erb.cs.uni-saarland.de/
- Consent form needs to inform participants about risks and their rights
  - in this seminar, we will send you a template, you need to fill in the details of your study
- Data needs to be stored securely (not publicly accessible) and in anonymized form
  - push results to Git, but only in anonymized form ("P12" instead of "John Smith")
  - be aware of GDPR (aka. DSGVO)

Part III: Implementing the Study in Unity



Best Practices in Unity-Based Studies



Clean Architecture → separate logic for:

- user interface (for you, the experimenter!)
- experiment control
- data logging

Plan for Flexibility and Debugging → allows for:

- quick adaptations during preparation
- thorough testing before data collection
- re-starts during the experiment without data loss

Core Functionalities to Include

Condition
(Independent Variables)

Participant	SubExperiment	Condition	ConditionIndex	MaterialIndex	MaterialName
P0	MATERIAL	CLOSED	0	0	Lightweight Plastic
P0	MATERIAL	CLOSED	1	0	Lightweight Plastic
P0	MATERIAL	CLOSED	2	0	Lightweight Plastic
P0	MATERIAL	HALF	3	1	Medium-Heavy Wood
P0	MATERIAL	FULL	4	2	Heavy Metal
P0	MATERIAL	HALF	5	1	Medium-Heavy Wood
P0	MATERIAL	FULL	6	0	Lightweight Plastic
P0	MATERIAL	HALF	7	1	Medium-Heavy Wood
P0	MATERIAL	FULL	8	2	Heavy Metal
P1	MATERIAL	CLOSED	0	0	Lightweight Plastic
P1	MATERIAL	HALF	1	1	Medium-Heavy Wood
P1	MATERIAL	CLOSED	2	0	Lightweight Plastic
P1	MATERIAL	HALF	3	1	Medium-Heavy Wood
P1	MATERIAL	CLOSED	4	0	Lightweight Plastic
P1	MATERIAL	HALF	5	2	Heavy Metal
P1	MATERIAL	FULL	6	2	Heavy Metal
P1	MATERIAL	FULL	7	1	Medium-Heavy Wood
P1	MATERIAL	FULL	8	0	Lightweight Plastic
P2	MATERIAL	HALF	0	0	Lightweight Plastic
P2	MATERIAL	HALF	1	0	Lightweight Plastic

Participant

Answers/Measurements
(Dependent Variables)

Data Logging

- First, think about:
  - what are your hypotheses?
  - what data do you need to investigate/prove it?
  - what tools will you use to analyze the data?
  - what (structured) format does these tools expect the data to be in?
- Then, implement:
  - automatic data logging during each trial
  - save data to file in a suitable format (e.g., CSV, JSON, text file) that aligns with the tools you use for analysis

Core Functionalities to Include

Participant	SubExperiment	Condition	ConditionIndex	MaterialIndex	MaterialName
P0	MATERIAL	CLOSED	0	0	Lightweight Plastic
P0	MATERIAL	CLOSED	1	0	Lightweight Plastic
P0	MATERIAL	CLOSED	2	0	Lightweight Plastic
P0	MATERIAL	HALF	3	1	Medium-Heavy Wood
P0	MATERIAL	FULL	4	2	Heavy Metal
P0	MATERIAL	HALF	5	1	Medium-Heavy Wood
P0	MATERIAL	FULL	6	0	Lightweight Plastic
P0	MATERIAL	HALF	7	1	Medium-Heavy Wood
P0	MATERIAL	FULL	8	2	Heavy Metal
P1	MATERIAL	CLOSED	0	0	Lightweight Plastic
P1	MATERIAL	HALF	1	1	Medium-Heavy Wood
P1	MATERIAL	CLOSED	2	0	Lightweight Plastic
P1	MATERIAL	HALF	3	1	Medium-Heavy Wood
P1	MATERIAL	CLOSED	4	0	Lightweight Plastic
P1	MATERIAL	HALF	5	2	Heavy Metal
P1	MATERIAL	FULL	6	2	Heavy Metal
P1	MATERIAL	FULL	7	1	Medium-Heavy Wood
P1	MATERIAL	FULL	8	0	Lightweight Plastic
P2	MATERIAL	HALF	0	0	Lightweight Plastic
P2	MATERIAL	HALF	1	0	Lightweight Plastic

Condition
(Independent Variables)

ppi	rial_num	start_time	end_time	trigger_type	igger_time	blink_time	response_time	noticed_trigger
P_1	6	534,8926	545,5526	BaseTrigger	540,1218	541,467	1,345215	False
P_1	7	546,3875	559,3864	FlashTrigger	551,4075	552,562	1,15448	True
P_1	8	560,677	571,7778	BaseTrigger	565,9669	568,4885	2,521606	False
P_1	9	573,019	582,8458	BaseTrigger	577,7945	579,3943	1,599792	False
P_1	10	584,1813	595,438	BaseTrigger	589,0552	590,7443	1,689087	False
P_1	11	596,4537	607,4265	AirPuffTrigger	601,6815	602,8899	1,208374	True
P_1	12	608,188	622,4482	BaseTrigger	616,4547	617,6847	1,230042	False
P_1	13	623,5198	632,6694	BlurTrigger	628,4714	629,3578	0,886414	False
P_1	14	633,4771	643,3768	BlurTrigger	638,7525	639,6631	0,910645	False
P_1	15	644,0867	655,0222	AirPuffTrigger	649,6065	652,3581	2,751587	True
P_1	16	656,0103	667,1113	BaseTrigger	661,5576	664,1548	2,597229	False
P_1	17	667,9877	679,1179	FlashTrigger	673,3292	676,0502	2,720947	True
P_1	18	680,1149	689,7615	BaseTrigger	685,2229	686,1405	0,917603	False
P_1	19	690,7535	728,4159	BlurTrigger	696,7505	701,8932	5,142761	True
P_1	20	729,7261	740,0731	BaseTrigger	735,1291	736,2335	1,10437	False
P_1	21	767,7072	776,611	AirPuffTrigger	772,5175	774,0389	1,521423	True
P_1	22	777,4922	787,9174	BaseTrigger	783,4012	784,5982	1,19696	False
P_1	23	788,8762	797,8083	ApproachingObjectTrigger	793,8524	794,2613	0,408875	True

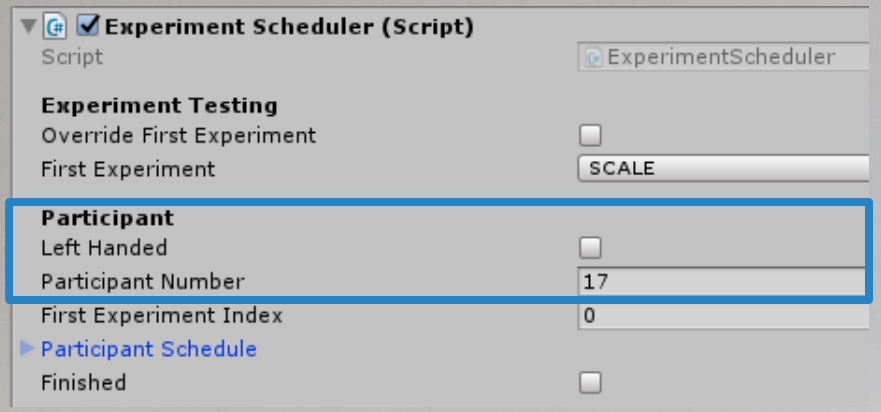
Participant

Answers/Measurements
(Dependent Variables)

Session Tracking

- Track all important parameters & results:
  - participant ID, participant properties
  - condition ID, condition name, condition settings
  - trial start time, trial end time, trial duration
  - dependent variables (i.e., answers & measurements)
  - completion status

Core Functionalities to Include



Configurable Parameters

- Make it easy to configure:
  - participant ID
  - important participant properties (e.g., handedness)
  - condition order
  - results directory
- Prevent errors:
  - validate important input: e.g., participant IDs
  - prevent unintended data loss due to file overwriting

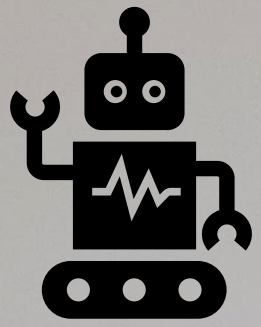
Core Functionalities to Include

ppid	trial_num	start_time	end_time	trigger_type	trigger_time	blink_time	response_time	noticed_trigger
P_1	6	534,8926	545,5526	BaseTrigger	540,1218	541,467	1,345215	False
P_1	7	546,3875	559,3864	FlashTrigger	551,4075	552,562	1,15448	True
P_1	8	560,677	571,7778	BaseTrigger	565,9669	568,4885	2,521606	False
P_1	9	573,019	582,8458	BaseTrigger	577,7945	579,3943	1,599792	False
P_1	10	584,1813	595,438	BaseTrigger	589,0552	590,7443	1,689087	False
P_1	11	596,4537	607,4265	AirPuffTrigger	601,6815	602,8899	1,208374	True
P_1	12	608,188	622,4482	BaseTrigger	616,4547	617,6847	1,230042	False
P_1	13	623,5198	632,6694	BlurTrigger	628,4714	629,3578	0,886414	False
P_1	14	633,4771	643,3768	BlurTrigger	638,7525	639,6631	0,910645	False
P_1	15	644,0867	655,0222	AirPuffTrigger	649,6065	652,3581	2,751587	True
P_1	16	656,0103	667,1113	BaseTrigger	661,5576	664,1548	2,597229	False
P_1	17	667,9877	679,1179	FlashTrigger	673,3292	676,0502	2,720947	True
P_1	18	680,1149	689,7615	BaseTrigger	685,2229	686,1405	0,917603	False
P_1	19	690,7535	728,4159	BlurTrigger	696,7505	701,8932	5,142761	True
P_1	20	729,7261	740,0731	BaseTrigger	735,1291	736,2335	1,10437	False
P_1	21	767,7072	776,611	AirPuffTrigger	772,5175	774,0389	1,521423	True
P_1	22	777,4922	787,9174	BaseTrigger	783,4012	784,5982	1,19696	False
P_1	23	788,8762	797,8083	ApproachingObjectTrigger	793,8524	794,2613	0,408875	True

Pay Attention to Details

- Test the data logging:
  - are floating point numbers correctly represented? (“,” vs. “.”)
  - is the delimiter correct? (“;” vs. “,”)
  - are the units correct? (“s” vs. “ms”; “deg” vs. “rad”; etc.)
  - are headers meaningfully chosen? (“angle2” vs. “angle_knee”)

Counterbalancing in Code



Automate as much as possible

- order of conditions based on counterbalancing
- setup of the scene based on condition
- data collection and saving

```
string[][] latinSquare = new string[][]
{
    new string[] { "Low", "Medium", "High" },
    new string[] { "Medium", "High", "Low" },
    new string[] { "High", "Low", "Medium" },
    new string[] { "High", "Medium", "Low" },
    new string[] { "Low", "High", "Medium" },
    new string[] { "Medium", "Low", "High" }
};
```

Latin Square from:
https://statpages.info/latinsq.html

```
int participantID = GetParticipantID();
int sequenceIndex = participantID % latinSquare.Length;
string[] conditionOrder = latinSquare[sequenceIndex];

// Example: Load each condition in order
foreach (string condition in conditionOrder)
{
    LoadThreatLevelCondition(condition);
    RunTrial();  // This would be your core VR task
}
```

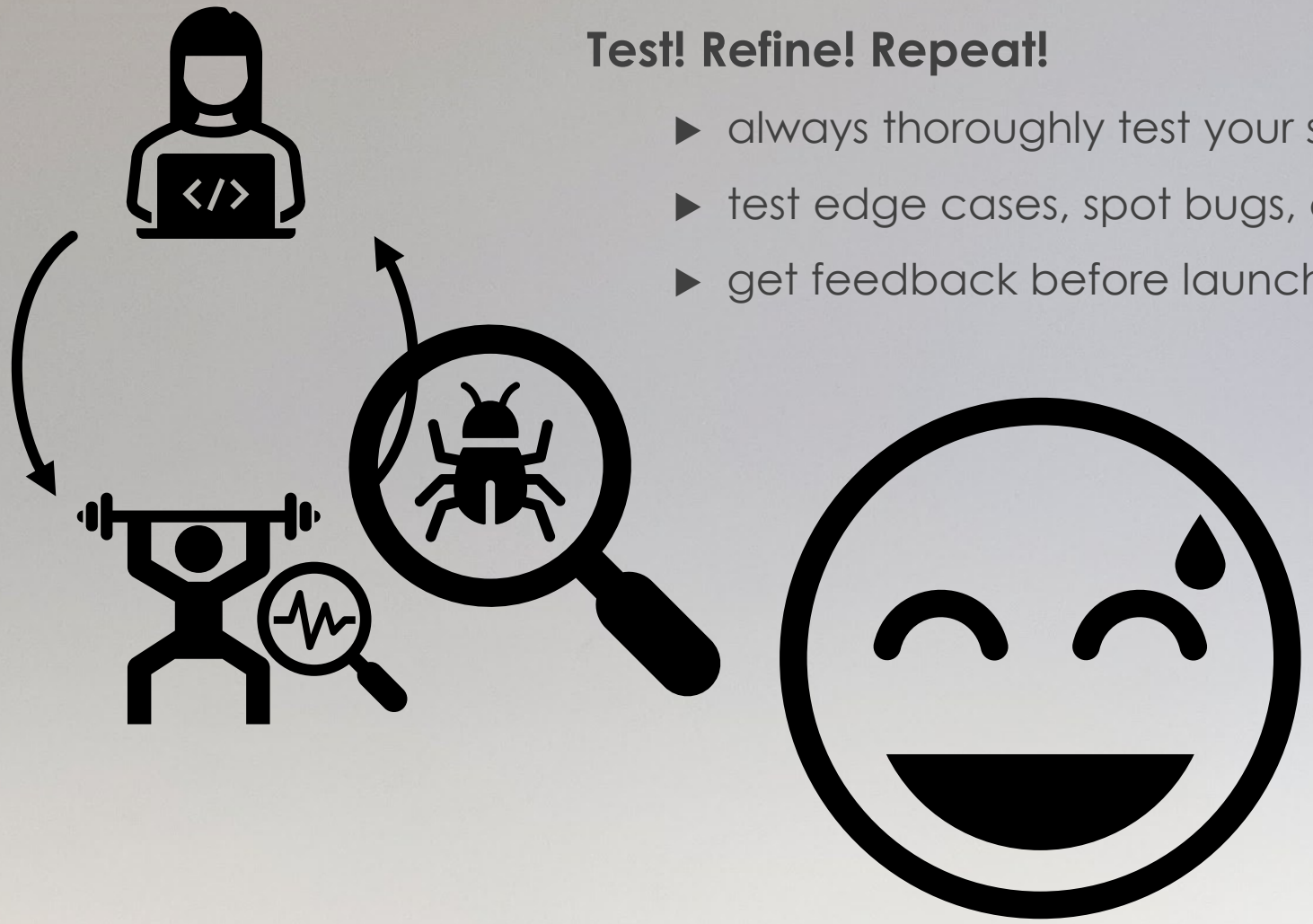
Example:

Participant 13

13 % 6 = 1

[M, H, L]

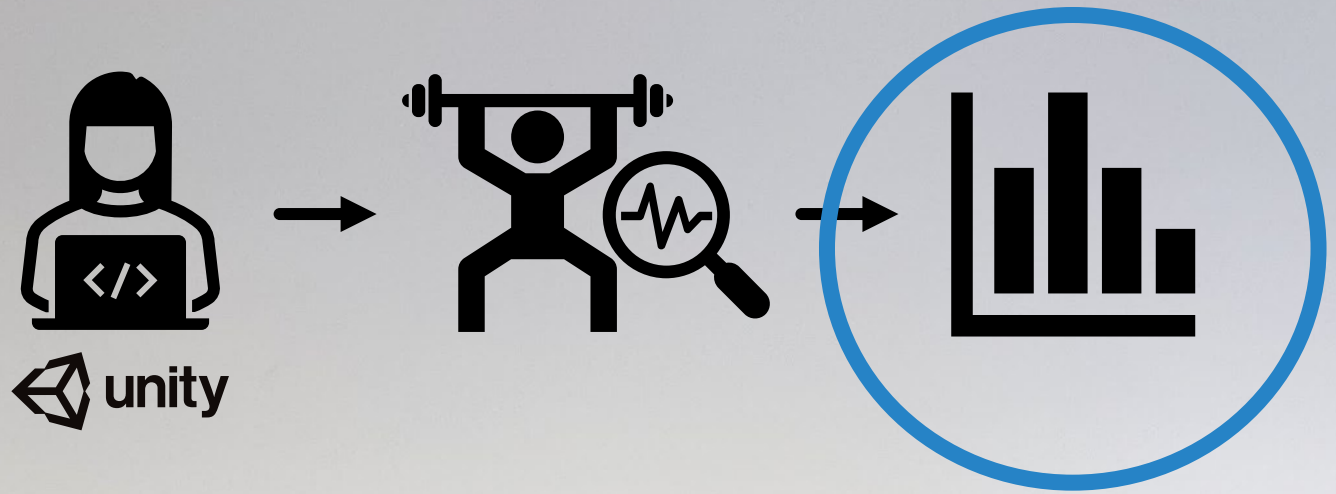
Pilot Testing



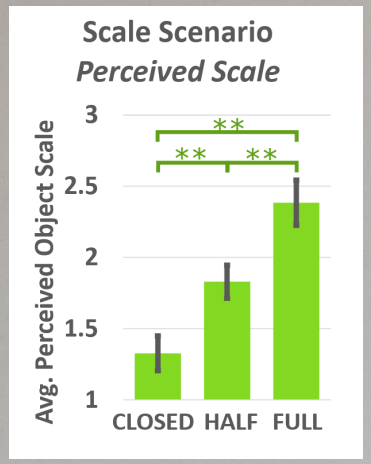
Test! Refine! Repeat!

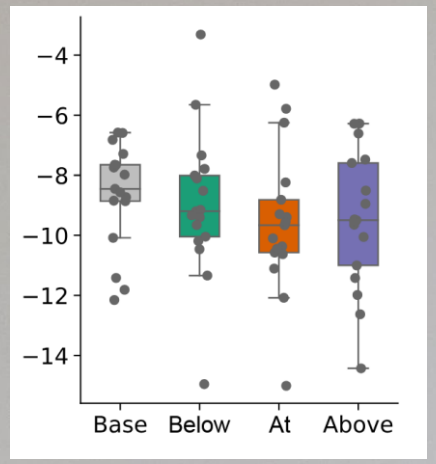
- always thoroughly test your setup
- test edge cases, spot bugs, confusing UI, data issues
- get feedback before launching the real study

Part IV: Analyzing & Reporting Results



Basic Data Analysis

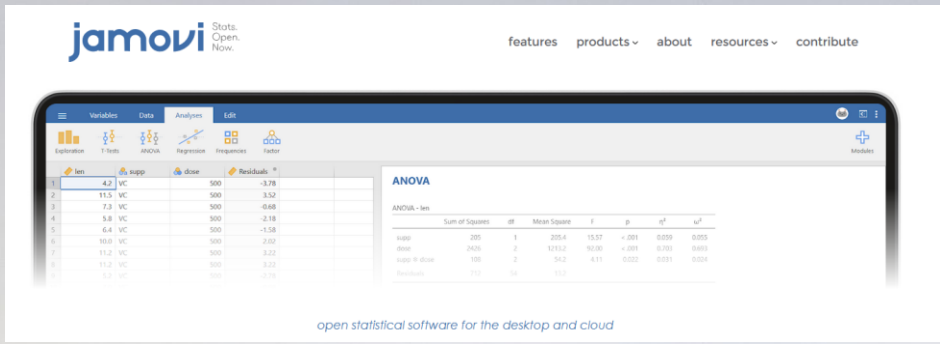


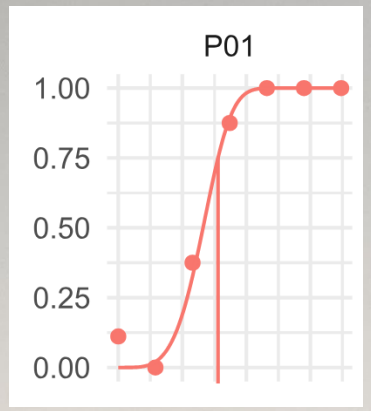


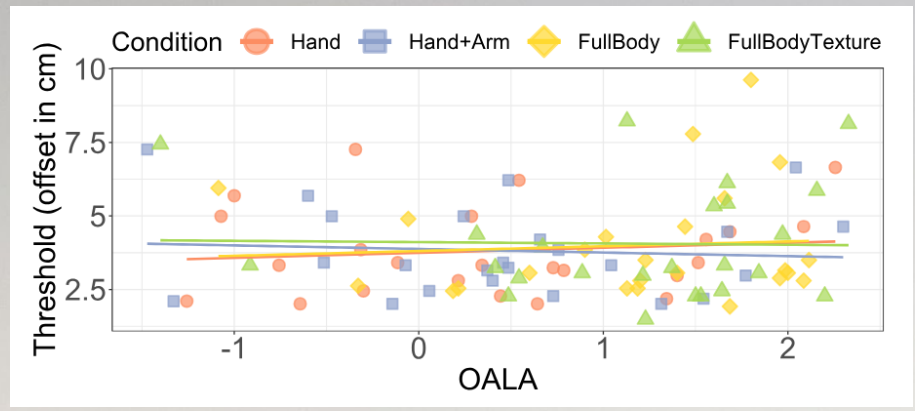
Plot your Data

- visualize it using suitable charts:
  - box plots
  - bar charts
  - line plots
  - scatter plots
  - …

Recommendation:
jamovi
(open-source statistics software)
https://www.jamovi.org/





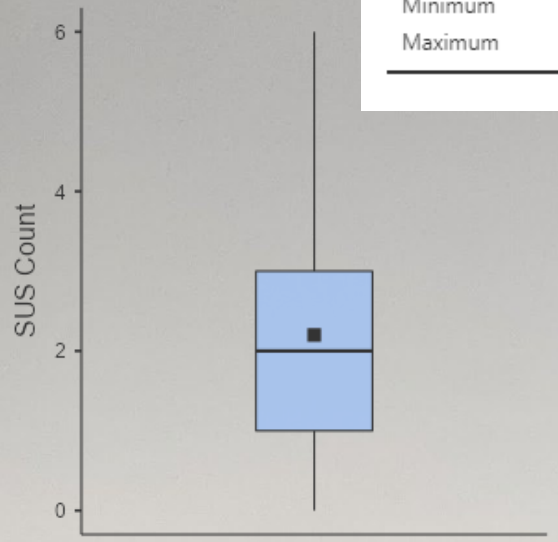


Basic Data Analysis

Deskriptivstatistik

Deskriptivstatistik

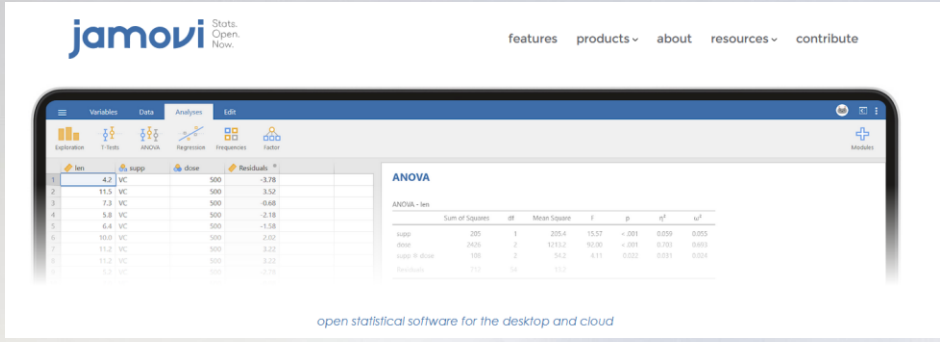
	SUS Count
N	25
Fehlend	0
Mittelwert	2.20
Median	2
Standardabweichung	1.78
Minimum	0
Maximum	6

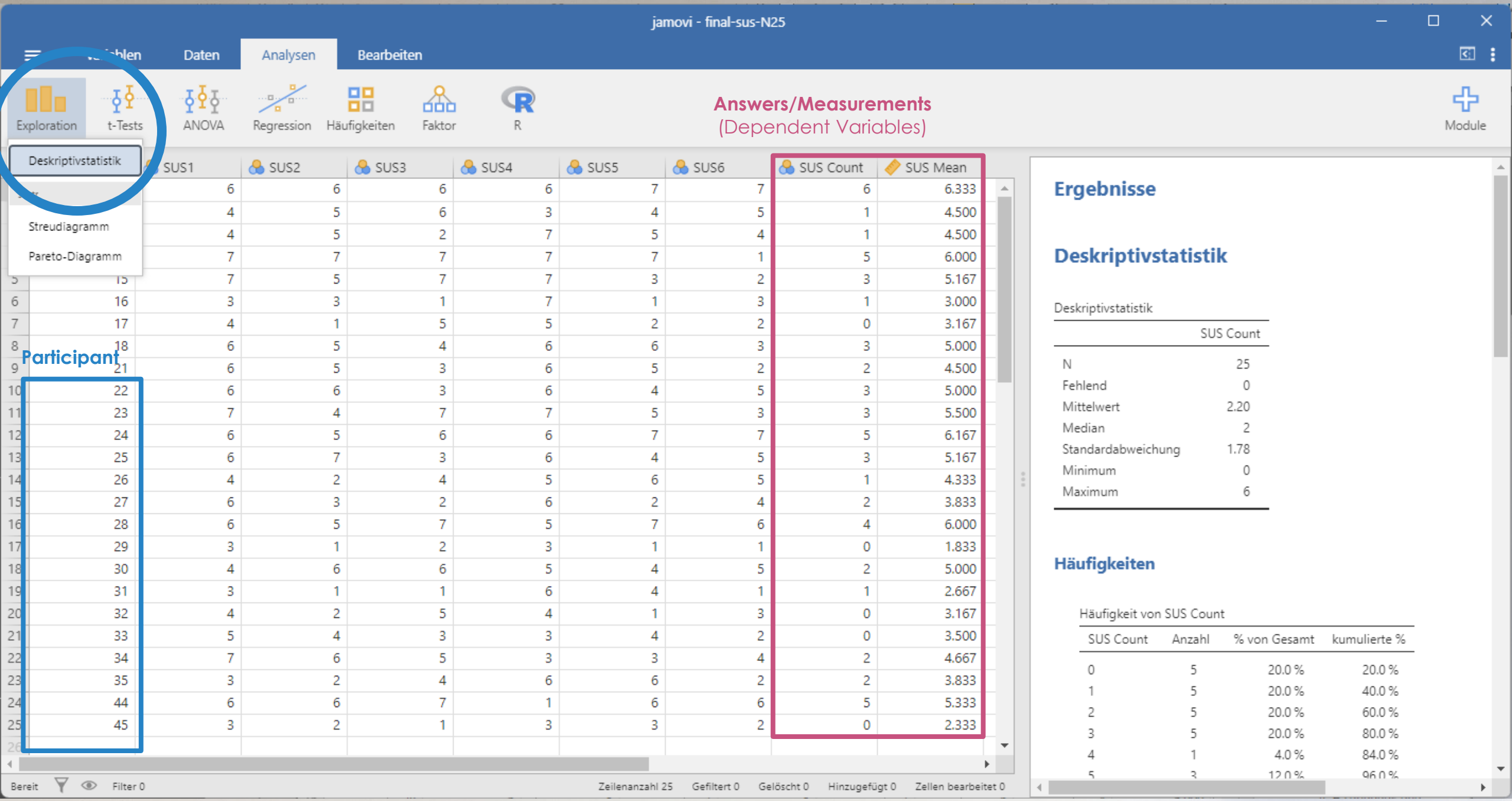


Compute Descriptive Statistics

- key statistical indicators that describe your data:
  - mean, median
  - min, max
  - standard deviation
- compute this for each condition
- plot comparison of conditions

Recommendation:
jamovi
(open-source statistics software)
https://www.jamovi.org/




jamovi - final-sus-N25
Daten
Analysen
Bearbeiten
Exploration
t-Tests
ANOVA
Regression
Häufigkeiten
Faktor
R
Module
Deskriptivstatistik
Streudiagramm
Pareto-Diagramm
Answers/Measurements
(Dependent Variables)
Participant
SUS1
SUS2
SUS3
SUS4
SUS5
SUS6
SUS Count
SUS Mean
Ergebnisse
Deskriptivstatistik
Deskriptivstatistik
SUS Count
N 25
Fehlend 0
Mittelwert 2.20
Median 2
Standardabweichung 1.78
Minimum 0
Maximum 6
Häufigkeiten
Häufigkeit von SUS Count
SUS Count | Anzahl | % von Gesamt | kumulierte %
0 | 5 | 20.0 % | 20.0 %
1 | 5 | 20.0 % | 40.0 %
2 | 5 | 20.0 % | 60.0 %
3 | 5 | 20.0 % | 80.0 %
4 | 1 | 4.0 % | 84.0 %
Bereit
Filter 0
Zeilenanzahl 25
Gefiltert 0
Gelöscht 0
Hinzugefügt 0
Zellen bearbeitet 0

jamovi - final-sus-N25

Variablen | Daten | Analysen | Bearbeiten

Exploration | t-Tests | ANOVA | Regression | Häufigkeiten | Faktor | R | Module

	Participant	SUS1	SUS2	S
1	10	6	6	
2	12	4	5	
3	13	4	5	
4	14	7	7	
5	15	7	5	
6	16	3	3	
7	17	4	1	
8	18	6	5	
9	21	6	5	
10	22	6	6	
11	23	7	4	
12	24	6	5	
13	25	6	7	
14	26	4	2	
15	27	6	3	
16	28	6	5	
17	29	3	1	
18	30	4	6	
19	31	3	1	
20	32	4	2	
21	33	5	4	
22	34	7	6	
23	35	3	2	
24	44	6	6	
25	45	3	2	
26				

Bereit | Filter 0 | Zeil... | G... | G... | Hin... | Zelle...

Deskriptivstatistik

Answers/Measurements
(Dependent Variables)

Participant
SUS1
SUS2
SUS3
SUS4
SUS5
SUS6
SUS Mean

Variablen: SUS Count

Aufgeteilt nach

Deskriptivstatistiken: Variablen über Spalten hinweg

Häufigkeitstabellen

Statistiken

Diagramme

Histogramme
- [x] Histogramm
- [ ] Dichte

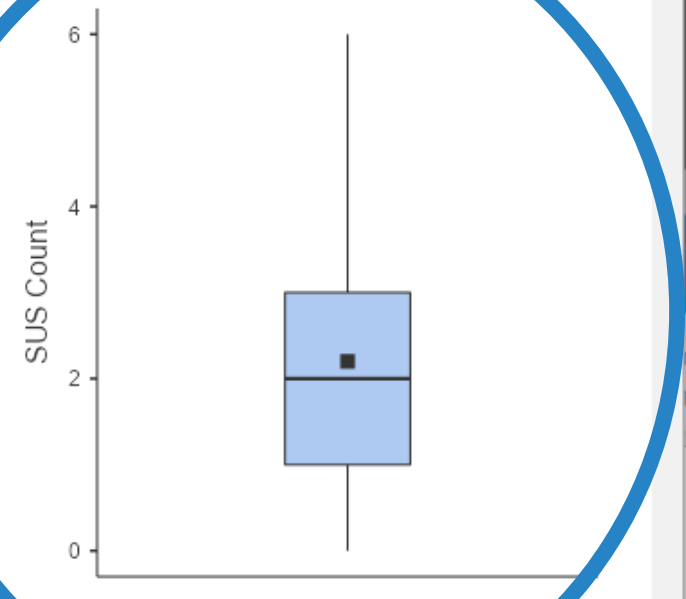
Box-Plots
- [x] Box-Plot
- [x] Ausreißer kennzeichnen
- [ ] Violin-Plot
- [ ] Daten

seitlich verschoben (jittered)

- [x] Mittelwert

Balkendiagramme
- [ ] Balkendiagramm

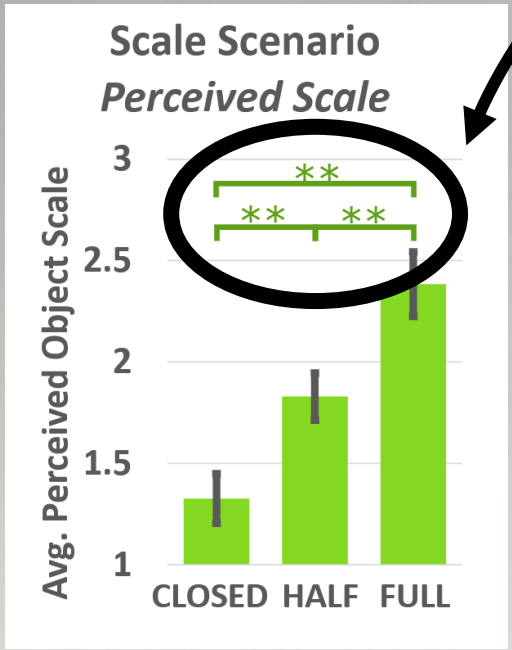
Q-Q-Diagramme
- [ ] Q-Q



Referenzen

Slightly More Advanced Data Analysis

https://www.jamovi.org/



Perform Statistical Tests (optional)

- statistical tests allow you to draw conclusions
  - they test, for example, if differences between conditions exist ("are statistically significant")
- common tests are:
  - ANOVA or Friedman test
  - t-test or Wilcoxon signed-rank test
- choose suitable tests based on:
  - experiment design (e.g., within-subjects vs. between-subjects)
  - properties of the collected data (are test assumptions (e.g., normality) met?)

Reporting Your Study

Final Report

The written final report should be 3-5 pages long (excluding references) and submitted as a PDF in ACM SIG format (2-column). The following LaTeX template should be forked into your own group:
https://umtl-git.dfki.de/seminar-sweat-and-survive/templates/final-report

A final report could be organized like this (*please also use figures/sketches in the document*):

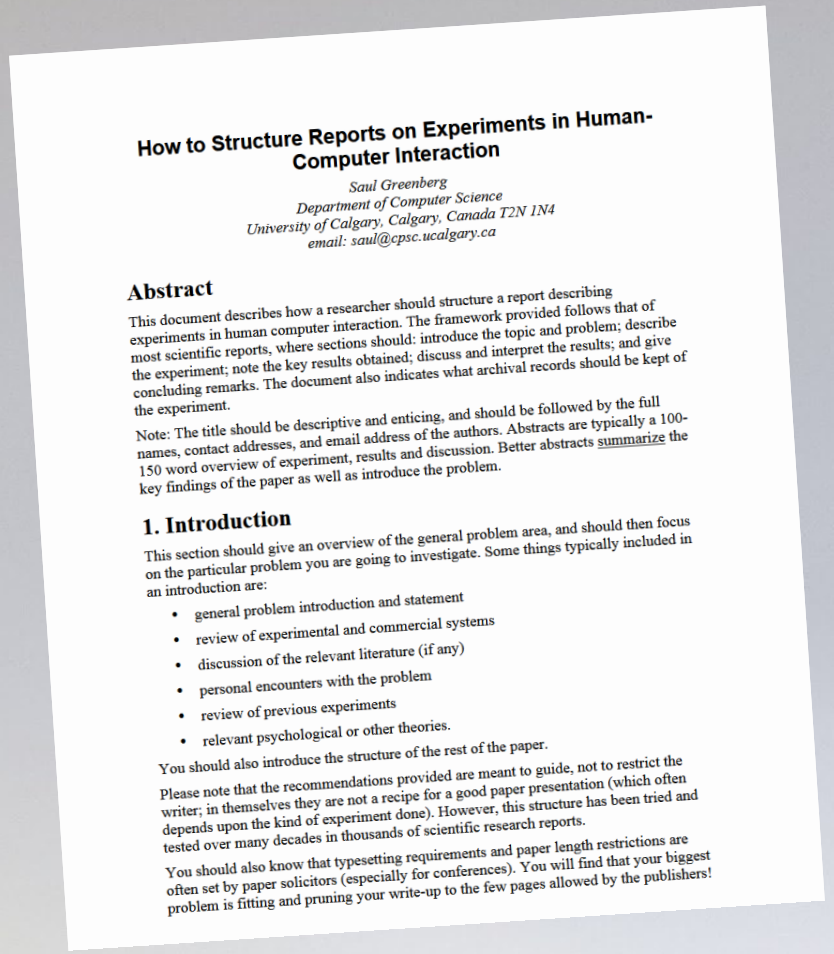
1. ***Abstract***
    - *Very short summary of your project and results.*

1. **Introduction**
    - Brief overview of the project and its objectives.
    - Summary of the initial concept and motivation.
2. **Related Work**
    - Review of relevant literature and previous work in VR sports applications and virtual threat scenarios.
    - Citation of at least 5 scientific papers related to the topic.
3. **Concept and Implementation**
    - Detailed description of the proposed VR application.
    - Explanation of how the application allows users to perform sports exercises while experiencing high adrenaline in virtual threat scenarios.
    - Technical details of the implementation, including software and hardware used.
4. **User Study**
    - Methodology of the user study conducted with the prototype.
    - Description of the participants and the procedures followed.
    - Data collection methods and tools used.

Reporting Your Study

General Structure:

1. **Introduction**
   (What are you testing and why?)
2. **Method**
   (Participants, Design, Tasks, Procedure)
3. **Results**
   (Plots, Statistics)
4. **Discussion**
   (What does the results mean?)
5. **Conclusion & Limitations**



How to Structure Reports on Experiments in Human-Computer Interaction

Saul Greenberg
Department of Computer Science
University of Calgary, Calgary, Canada T2N 1N4
email: saul@cpsc.ucalgary.ca

Abstract

This document describes how a researcher should structure a report describing experiments in human computer interaction. The framework provided follows that of most scientific reports, where sections should: introduce the topic and problem; describe the experiment; note the key results obtained; discuss and interpret the results; and give concluding remarks. The document also indicates what archival records should be kept of the experiment.

Note: The title should be descriptive and enticing, and should be followed by the full names, contact addresses, and email address of the authors. Abstracts are typically a 100-150 word overview of experiment, results and discussion. Better abstracts summarize the key findings of the paper as well as introduce the problem.

1. Introduction

This section should give an overview of the general problem area, and should then focus on the particular problem you are going to investigate. Some things typically included in an introduction are:

- general problem introduction and statement
- review of experimental and commercial systems
- discussion of the relevant literature (if any)
- personal encounters with the problem
- review of previous experiments
- relevant psychological or other theories.

You should also introduce the structure of the rest of the paper.

Please note that the recommendations provided are meant to guide, not to restrict the writer; in themselves they are not a recipe for a good paper presentation (which often depends upon the kind of experiment done). However, this structure has been tried and tested over many decades in thousands of scientific research reports.

You should also know that typesetting requirements and paper length restrictions are often set by paper solicitors (especially for conferences). You will find that your biggest problem is fitting and pruning your write-up to the few pages allowed by the publishers!

Recommended Read!

http://csci.viu.ca/~carruths/Courses/CSCI310/Readings/Greenberg_writing_a_report.pdf

Wrap-Up & Resources

Tips for Success

Don’t over-complicate

Respect participant effort

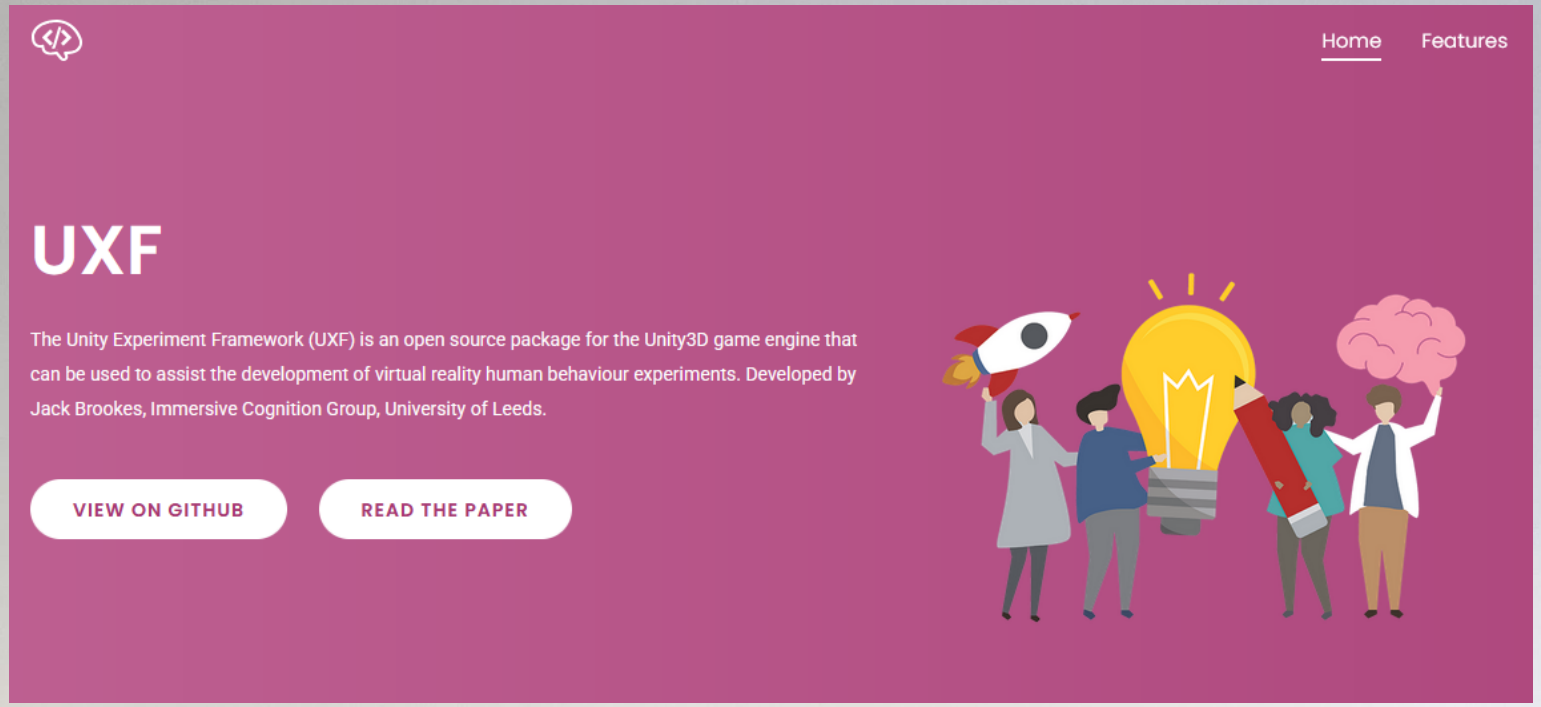
Test everything beforehand

Document each step
as if someone else needs to replicate it

Useful Unity Frameworks

The Unity Experiment Framework

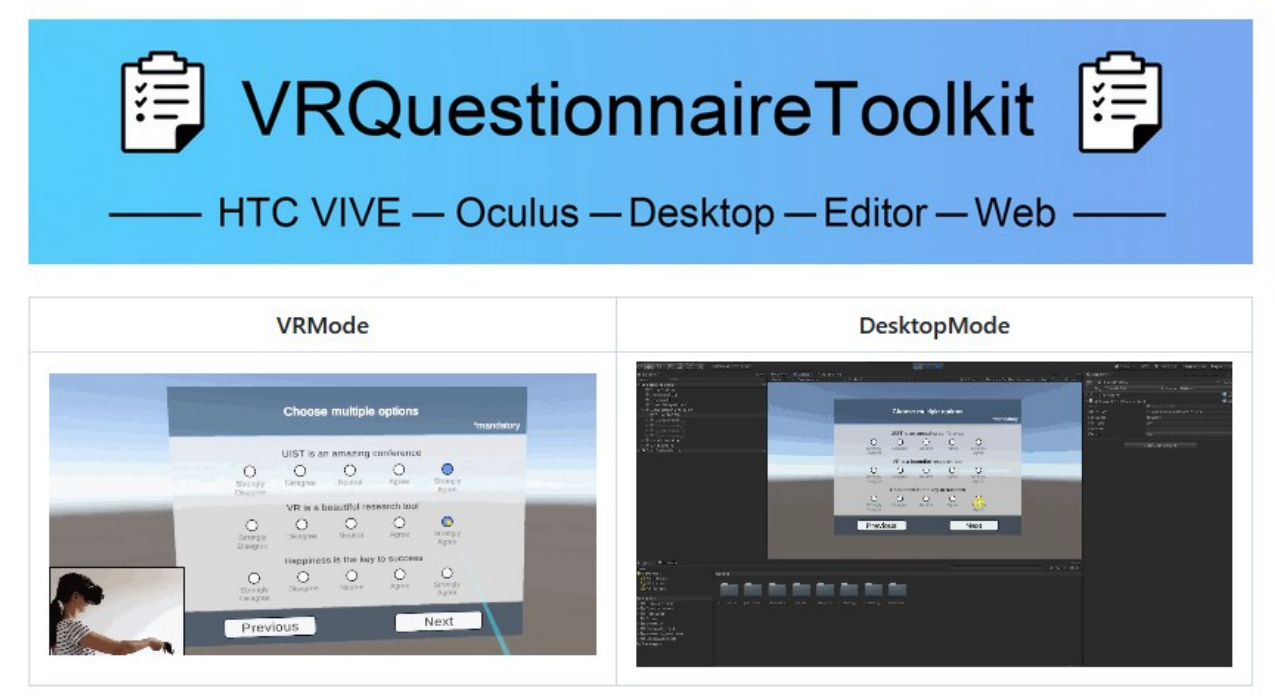
https://immersivecognition.github.io/unity-experiment-framework/



Useful Unity Frameworks

The VR Questionnaire Toolkit

https://github.com/MartinFk/VRQuestionnaireToolkit

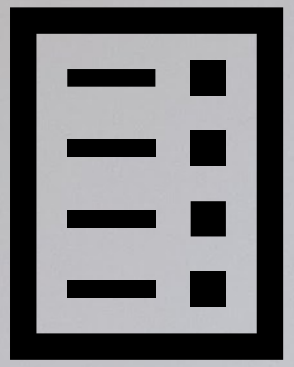


Important!

Document your VR application and study with videos!

Include videos in your final presentation!

Important!



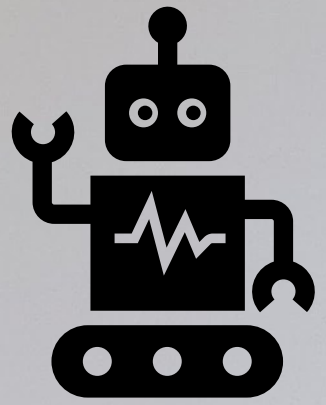
Include the following questionnaires in your study:

Slater-Usoh-Steed (SUS) Presence Questionnaire (after each condition)

Borg Rating of Perceived Exertion (RPE) Scale (after each condition)

… potentially 1 or 2 more questionnaires (we will inform you)

Important!



If using Gen-AI for the Final Report, add a statement how you used it in the Acknowledgements section!

Gen-AI is only a tool – not a replacement for your own work.

Using Gen-AI content without acknowledging it can be plagiarism!

Q&A and Discussion

Bachelor/Master Thesis Opportunity

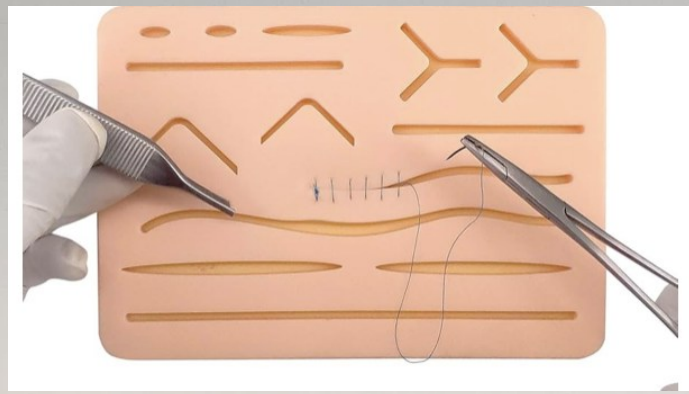
Topic 1: Anticipating User Intents in Natural Human-Robot Collaboration using Multimodal Interaction

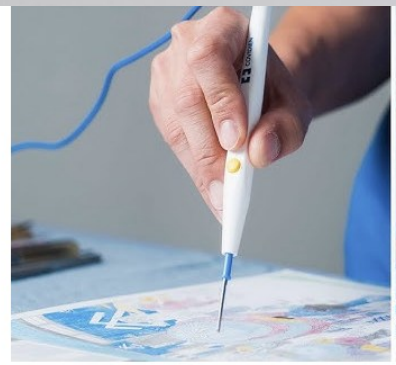
- **User intent recognition** using **EEG, gestures, and speech in realistic industrial tasks**. This project moves beyond controlled lab settings to enable natural, self-paced interaction with robots.
- You will work with real-world multimodal data and develop methods for intent prediction, signal fusion, and adaptive HRC systems.
- Students interested in BCI, HRI, multimodal AI, Machine learning, real time systems.

Contact Details:
mansi.sharma@dfki.de

Topic 2: Predicting Perceived Task Difficulty in Surgical Training Using EEG and Eye Tracking

- **EEG and eye tracking to objectively assess factors that contribute to perceived task difficulty**, such as time pressure, use of the non-dominant hand, in surgical practice tasks like suturing
- You will work with biometric data, perform feature extraction and train a baseline machine learning model to predict whether a task is perceived as difficult or not based on EEG and eye tracking signals.
- Students interested in EEG, eye tracking, and machine learning.





Contact Details:
mansi.sharma@dfki.de