Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

IARIA

Wolfgang Leister

Keynote

Assessing and measuring human factors – Trends and Developments

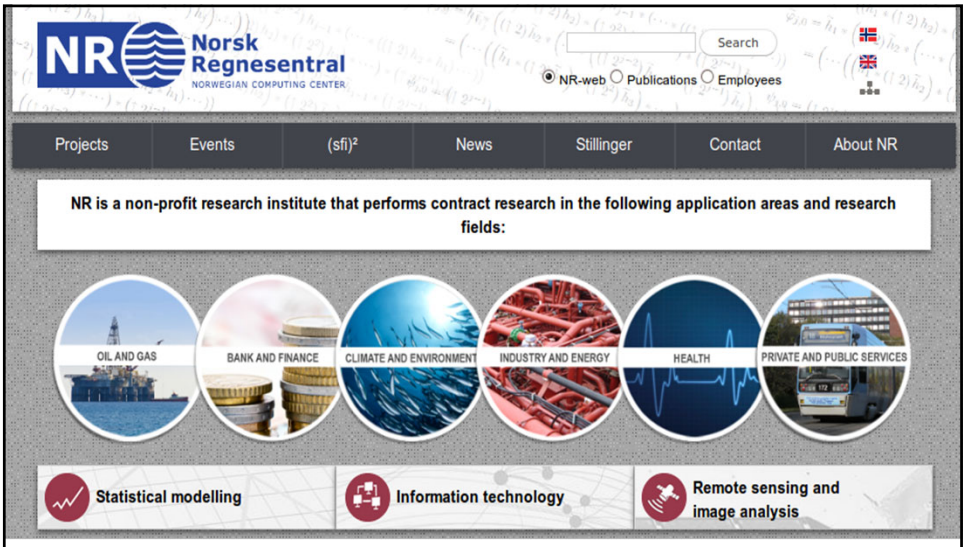
- PESARO 2022
- Barcelona - April 2022 (virtually from Oslo)

1

Dr. rer.nat. Wolfgang Leister

- Acting Research Director, Norsk Regnesentral, Oslo, Norway.
- Studied informatics at the University of Karlsruhe (now KIT). Dissertation on a subject related to computer graphics and user interfaces (1991).
- Industry experience;
- Joined Norsk Regnesentral in 1997.
- Leads the Department of Applied Research in Information Technology (DART).
- Applied Research in the following areas:
  - Digital Inclusion
  - Digital Security
  - Digital Transformation

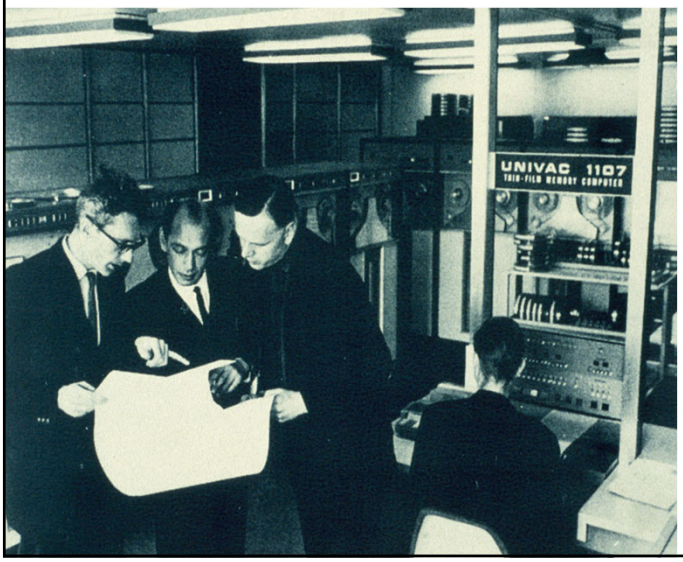
3

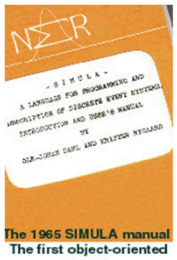


Norsk Regnesentral works with applied research. Established 1952. Ca 90 researchers. Customers are private companies, the Research Council of Norway, EU, public sector, international companies, ...

4

Object orientation and SIMULA





The 1965 SIMULA manual
The first object-oriented report

5

Outline of the talk

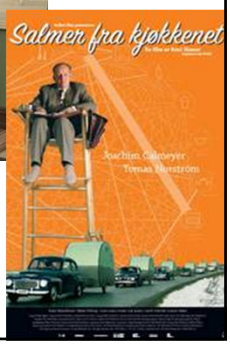
How can we measure people's engagement with a program, a museum exhibition, and a hike through nature?

The talk will be about assessing and measuring human factors using different methods in different contexts and for various user groups.

What impact does the pandemic have on this field?

6

How to observe and measure engagement and behaviour?
Reference to the movie «Kitchen stories» by Bent Hamer (2003).
Can we replace Folke Nilsson by a Kinect ?



7

Observations – Use of video footage – Measuring outcome.
Quantitative vs. Qualitative? Or both?

8

To evaluate user behaviour, we need to observe the visitors and find suitable metrics that describe installations, engagement, and other qualities.

-

eQuiz @ Engineerium

9

Human Factors

- Human factors and ergonomics is the application of psychological and physiological principles to the engineering and design of products, processes, and systems. The goal of human factors is to reduce human error, increase productivity, and enhance safety, system availability and comfort with a specific focus on the interaction between the human and the engineering system.[1] (Wickens, Gordon, Liu)

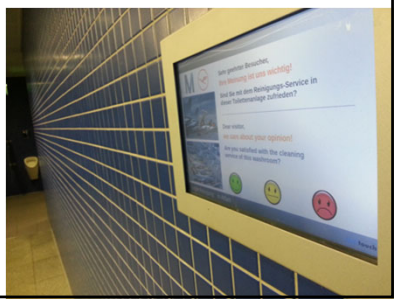
Increase user experience ?

10

Digital experiences can be assessed with objective and subjective measurements. But, would you answer a questionnaire after a museum visit with 30 questions about installations you even did not look at ?

Objective measurements
- QoS
- Network monitoring
- Image & audio metrics
- ...

Subjective measurements
- QoE
- Questionnaires
- Likert & similar
- Mood detection
- Sentiment analysis
- ...



11

How can we measure? We tracked visitors with a Kinect and measured how long a visitor stayed. (Forskerpool project with industry partner Expology)

How to count visitors in a museum?

- wear marks on carpet
- personnel count visitors
- cameras
- beacons

What are the challenges?

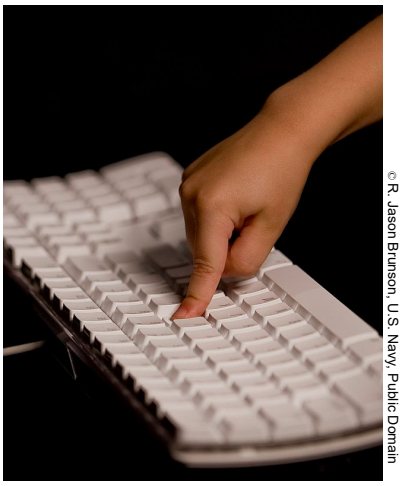
- disturbing the visitor impacts result
- logistics must be simple
- accuracy

What did we find?

- We could count activity
- Some technical issues
- Multiple entries are not detected
- We could observe cleaning personnel
- Certain events (evening with librarians, kids event, ...) and school class visiting were different.
- How to interpret data?
- Observe a person walking around installation using phone.
- Make sure that privacy issues are correctly addressed

12

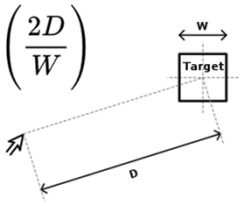
Measuring Human Factors



© R. Jason Brunson, U.S. Navy, Public Domain

Fitts' Law: (Paul Morris Fitts, 1954)

$$\mathrm{ID} = \log_2\left(\frac{2D}{W}\right)$$



Hick's Law: (William Edmund Hick / Ray Hyman, 1951

$$T = b \cdot \log_2(n+1)$$

13

SUS – System Usability Scale (Brooke 1996)

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

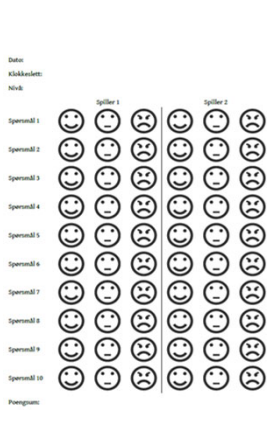
Participant	SUS Statement										Score	Grade
	1	2	3	4	5	6	7	8	9	10		
1	4	5	4	4	3	1	5	2	2	5	52.5	D
2	3	2	4	2	5	1	4	2	4	5	70	C
3	2	1	3	1	4	4	5	2	1	1	65	C
4	4	2	4	2	3	3	4	2	3	3	65	C
5	3	1	5	1	5	2	4	1	2	1	82.5	A
6	1	3	3	1	2	4	2	1	1	1	47.5	F
7	4	1	5	1	3	2	5	2	2	2	77.5	B+
8	1	1	3	1	1	3	4	3	1	3	47.5	F
9	1	3	3	2	4	2	4	1	1	2	57.5	D
10	4	2	4	4	3	2	3	2	3	4	57.5	D

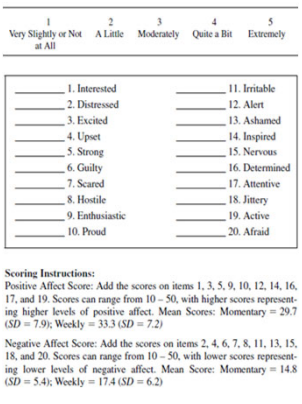
… the average score is 62.77 and a median score of 61.25, which corresponds to a C- and D respectively

(not unusual for a very first implementation of a prototype)

14

Manual methods for subjective and objective assessment are compared with the automatically retrieved assessment. Here are the valence tracker, the PANAS, and the questionnaire





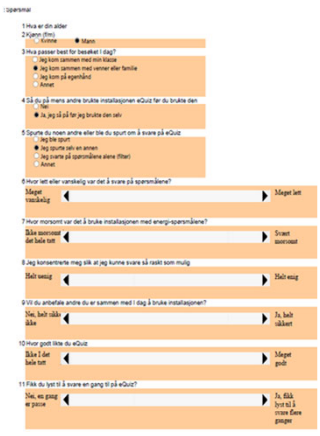
1 Very Slightly or Not at All | 2 A Little | 3 Moderately | 4 Quite a Bit | 5 Extremely

_____ 1. Interested	_____ 11. Irritable
_____ 2. Distressed	_____ 12. Alert
_____ 3. Excited	_____ 13. Ashamed
_____ 4. Upset	_____ 14. Inspired
_____ 5. Strong	_____ 15. Nervous
_____ 6. Guilty	_____ 16. Determined
_____ 7. Scared	_____ 17. Attentive
_____ 8. Hostile	_____ 18. Jittery
_____ 9. Enthusiastic	_____ 19. Active
_____ 10. Proud	_____ 20. Afraid

Scoring Instructions:

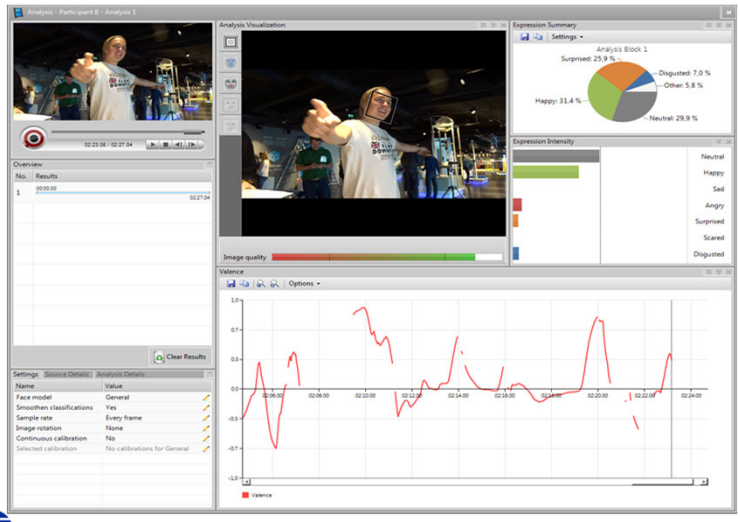
Positive Affect Score: Add the scores on items 1, 3, 5, 9, 10, 12, 14, 16, 17, and 19. Scores can range from 10 – 50, with higher scores representing higher levels of positive affect. Mean Scores: Momentary = 29.7 (*SD* = 7.9); Weekly = 33.3 (*SD* = 7.2)

Negative Affect Score: Add the scores on items 2, 4, 6, 7, 8, 11, 13, 15, 18, and 20. Scores can range from 10 – 50, with lower scores representing lower levels of negative affect. Mean Score: Momentary = 14.8 (*SD* = 5.4); Weekly = 17.4 (*SD* = 6.2)



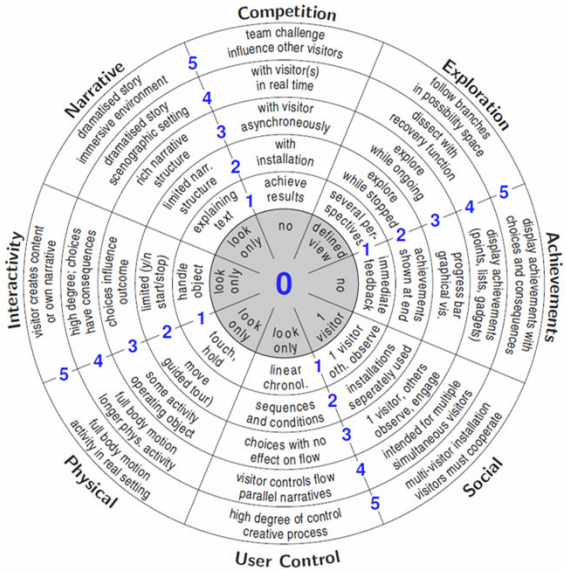
15

The FaceReader software can assess the dominant emotion.



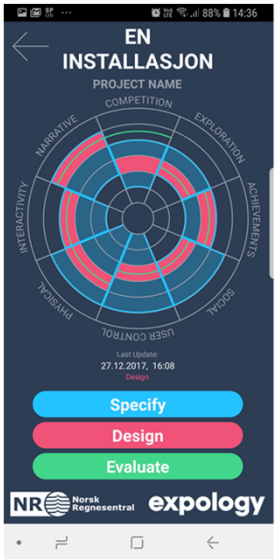
16

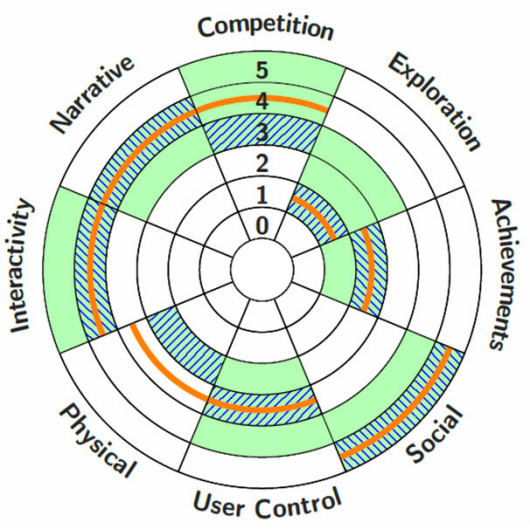
How could we characterise the exhibits so that we can get an idea what to change for better user experience?



19

Example of an evaluation of an exhibit





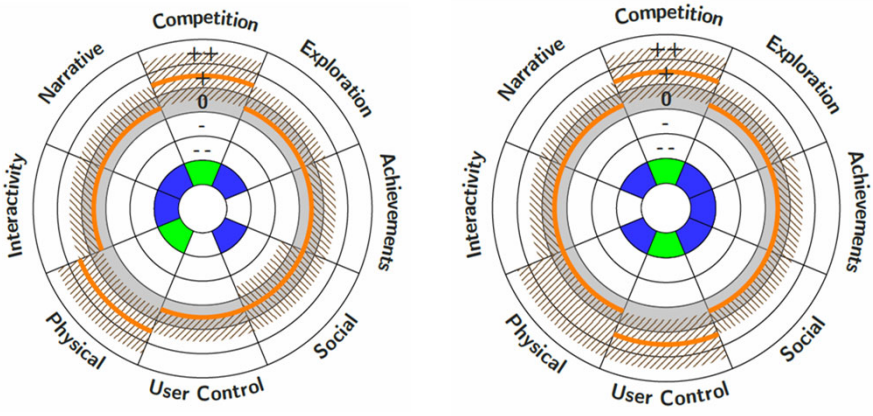
20

Questionnaires are part of the assessment with a «standardised» questionnaire of 8 questions + a few about like, recommend, play again, ...

For each of the eight dimensions of engagement, the following scale is used: −2 (much less), −1 (less), 0 (as now), 1 (more), 2 (much more).	
Q_C	Should there be more or less competition between groups and participants in the game?
Q_N	Should the storyline and roles in the game be more evident or less evident?
Q_I	Should there be more or less feedback on the choices you did in the game?
Q_P	Should there be more or less physical activity in the game?
Q_U	Should the story in the game be more or less influenced by what you did during the game?
Q_S	Should a considerable part of the game be done in a group or should there be more individual tasks to solve?
Q_A	Should there be more or less feedback on how well you are doing during the game?
Q_E	Should there be more or less possibilities to go in depth with extra content to solve the tasks in the game?

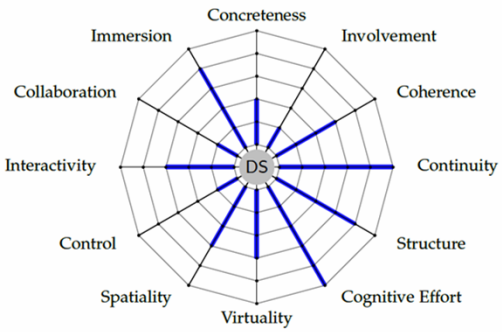
21

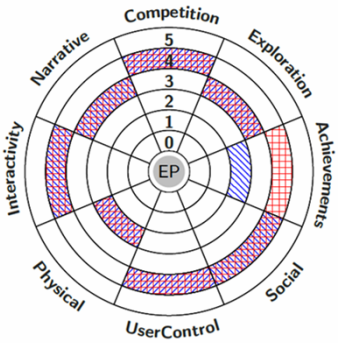
Change diagram for Engineering Challenge (Engineerium)
left: lower secondary school, right: higher secondary school



22

How to characterise the narrative? Engagement profile, dimension star, and GameFlow for Motorway of the Ocean

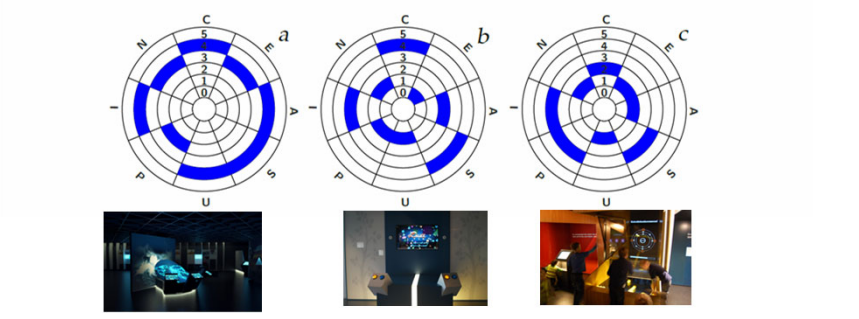


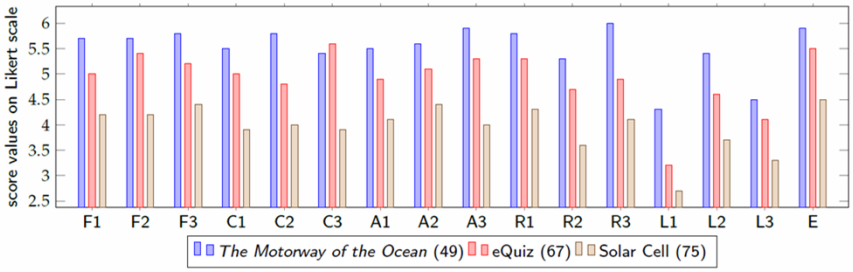


criteria	average	detailed values
concentration	4.3	(5, 5, 5, 3, 4, 4)
challenge	2.0	(3, 1, 1, 3)
player skills	4.6	(5, 5, 0, 0, 4, 4, 5)
control	2.8	(3, 4, 3, 0, 2, 2)
clear goals	5.0	(5, 5)
feedback	5.0	(5, 5, 5)
immersion	2.4	(3, 3, 2, 2, 2)
social interaction	3.0	(3, 0, 0)
average	3.6	

23

We compared three installations – Motorway of the Ocean scores highest.

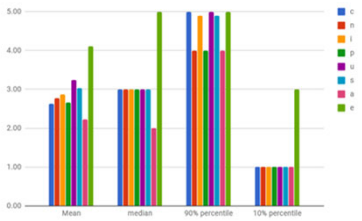


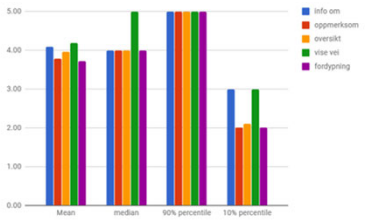


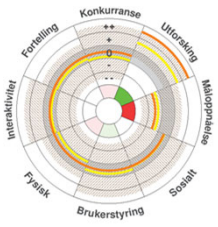
24

We applied this method to evaluate an app to guide pensioners to look at art in a public space (project with Larvik Kunstformidling).

- Is an app something the target group would like?
- We developed an interview guide.
- We used Google Forms for the questionnaires.
- 112 respondents
- Engagement profile: Explore, but not showing achievements.
- Functionality: show way directions.

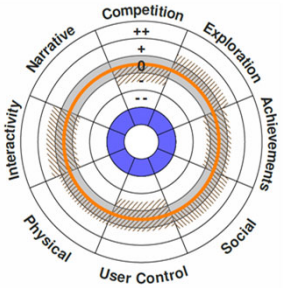


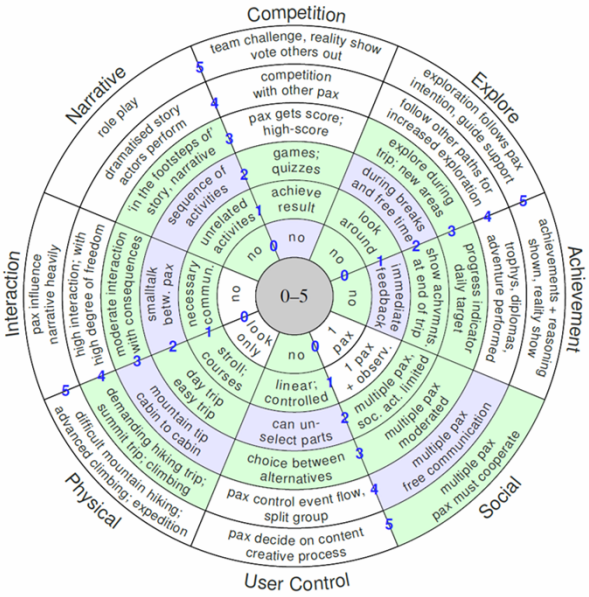




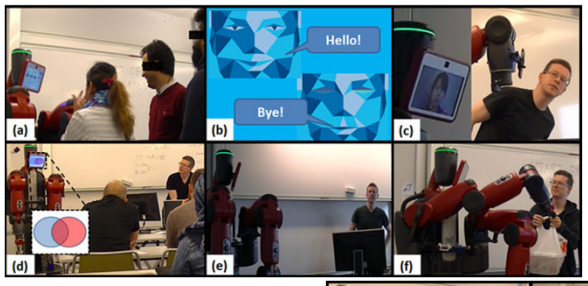
28

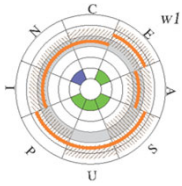
We applied the Engagement Profile to different areas; here: trekking arrangements.

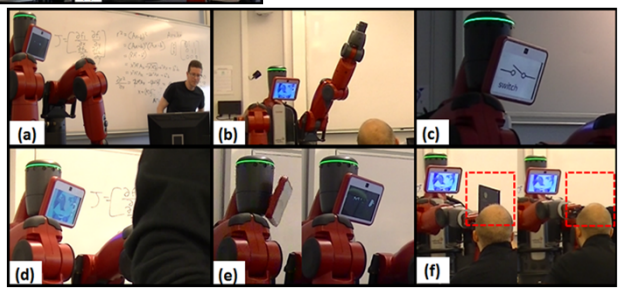




We used this methodology to evaluate the use of a robot in a teaching context. (Collaboration with Halmstad University, Sweden)







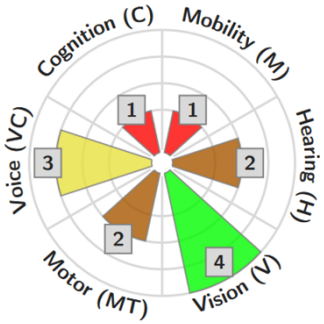
In museums and science centres, learning (experiences) must be universally designed, so that everybody can get an engaging experience.

- How can we measure universal design in museums and science centres?
- Informal learning, micro-learning
- Accessible for persons with physical and sensory impairments
- Multi-modal and multi-sensory access
- Accounting for diversity in cognition (impairments and normal)
- Creating engaging experiences for everybody

31

Human factors research considering persons with disabilities

- How well suited are exhibitions for persons with disabilities?
- Making exhibitions suitable for persons with disabilities will make these suitable for a wider audience!
- How to measure?
- WCAG (developed for web pages)
- What about using XR (VR/AR/MR) technologies for all? Is this possible?



© Maurizio Pesce, CC-BY

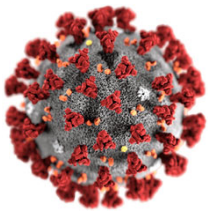
32

Observation: The Covid-19 Pandemic has changed research in human factors

HCI-related papers submitted lately report of challenges related to the pandemic.

Pandemic-related restrictions:

- Focus groups could not be performed as planned.
- Testing of artefacts was challenging.
- Methodology changes were necessary.
- Use of simulations, as artifacts could not be tried out.
- Projects were delayed → no time left for evaluations.

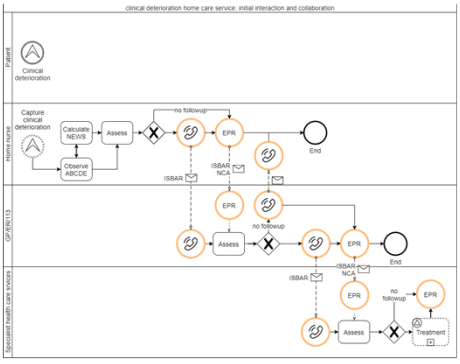


Are the replacement-methods good enough?

33

Example: Focus Group Workshops were changed to comply with restrictions.

Setting: Development of communication scenarios in healthcare communication under suboptimal conditions.



Focus group workshops using the customer journey method were "digitized" using a spreadsheet.

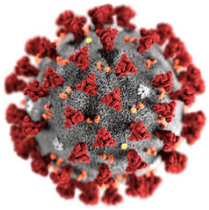
We planned for a physical workshop, with a virtual workshop as a possible backup-plan.

The template was shared on a screen, and participants also brought their own laptops, to access shared documents.

This method can be further developed to improve focus group workshops also post-pandemic.

34

Observations while reviewing papers



When testing artefacts:

- Use of video stimuli instead of a physical setup – Is the stimulus strong enough?
- Reduced test groups with only few persons – Statistical significance?
- Interpretation of observations – instead of measuring (change the discipline)
- Testing in a different setting or environment – Is the result still relevant?

35

Takeaway

- The pandemic has changed the field
  - – The “new” methods need to be evaluated …
- Quantitative methods need to be developed further
  - – and combined with qualitative methods …
- Methods considering a diversity of disabilities need to be developed.
  - – blind, visually impaired, hearing disability, motor, mental, …
- Methods considering a diversity of target groups need to be developed.
  - – children, elderly, technology-averse, professions, …
- Need methods with more flexibility to adapt to changed circumstances.

37

Thank you for your attention!

Questions?

Wolfgang Leister
Norsk Regnesentral
wolfgang.leister@nr.no

38