PANDERAM

Laboratory Study 1 PANDERAM Prototype App Analysis and Isomorphism

Results AP 7.2 and 7.3

Susen Döbelt & Dominik Lange
Chemnitz, April 2023

secuvera
BSI-zertifizierter IT-Sicherheitsdienstleister und Prüfstelle

DAI-Labor
TU Berlin

Fraunhofer
SIT

ALLGEMEINE &
PSYCHOLOGIE
HUMAN FACTORS

INSTITUT
FÜR
TECHNIK
UND
JOURNALISMUS

Overview

Introduction: Goals, organization, procedure, data preparation and evaluation

Results

- **Laboratory study Task 1 and 2**
- **Laboratory study Task 3: App isomorphism**

Summary

Outlook

Goal(*s*) of the survey

Investigation of the PANDERAM prototype with regard to potential improvements for **usability** and **user experience** as well as **fit** and **effectiveness** for different **behavior levels.**

Research Questions:

How do individuals of different behavioral levels evaluate the prototype that matches the behavioral level vs. the prototype that does not match?

Does the PANDERAM prototype increase self-efficacy in relation to data and privacy protection?

How do users rate the transparency of the PANDERAM prototype?

Can conclusions be drawn about the intention to use the behavior level, app variant and transparency?

Which app isomorphism method provides the better app alternative suggestions from the user's perspective?

Organization 1/2

Schedule

- Start of conception (05/2022)
- Implementation and testing questionnaires (07/2022)
- Start (recruitment): 11/13/2022
- Start (laboratory test): 12/07/2022
- Completion: 31/01/2023 (5 weeks of testing)

Organization 2/2

Recruitment

- Study call with link to the recruitment questionnaire
- **Compensation:** Completing recruitment: 0.5 test person hours or participation in a raffle of 5 x €20, remuneration for the laboratory experiment: 1.5 test person hours or €25 in cash)
- **Distribution:** via TUC study participation mailing list (several reminders), AHF website, notices in Wilhelm-Raabe-Str./TU campus/canteen/private environment, chat and Facebook groups private environment

N = 129 individuals completely filled out the recruitment questionnaire.

- Criteria for selection laboratory test:
  - Android operating system >6.0 (n = 70, = 54% from the recruitment questionnaire)
  - Behavior level assignment possible (n = 69) and balance of the behavior level groups
Ultimately, N = 38 individuals participated in the laboratory experiment.

Procedure 1/5

Before starting the investigation

- **Rejection** of persons from the recruitment questionnaire and **remuneration** of those who have registered
- Confirmation via e-mail to selected people
  - Request **date selection**
- **Assignment to the experimental condition** (prototype matching or not matching for the behavior level) and questionnaire variant A or B (randomized display of content)

Procedure 2/5

Investigation

- **Welcome** and request to read and sign the **participation and privacy information**
- Information available **equipment**: laptop and mouse for completing the questionnaire, test smartphone and PANDERAM prototypes installed on it
- **Notes** from the experimenter:
    - PANDERAM app to be evaluated is still a prototype (some areas are not functional or not fully developed yet)
    - Please do not delete any apps from the test smartphone
    - Follow the instructions in the questionnaire and keep an eye on the timer
    - If you have any questions, contact the experimenter

start of the experiment

- Collection of self-descriptive variables: Affinity for technology, Smartphone competence and Need for cognition
- Pre-measurement of self-efficacy (agreement of statements and rating)
- Free exploration of the PANDERAM prototype (10min)

Procedure 3/5

Investigation

Task 1:
Learn about the Freeletics app using the PANDERAM prototype.

- Describe the risk factors of this app in your own words (5min)

- Performance measurement pre : What is the current risk score of the app "Freeletics"?

Task 2:
Using the PANDERAM prototype, take measures to **minimize the risk** of the app "Freeletics".

- Describe the measures you have taken in your own words (5min)

- Performance measurement post: What is the current risk score of the app "Freeletics" now?

Procedure 4/5

Investigation

Evaluation PANDERAM prototype
Quantitative assessment by:

- UTAUT: Performance expectation, effort expectation, attitude towards technology
- meCUE 2.0: Usability, usefulness, intention to use, positive and negative emotions
- PET-USES: General usability, privacy preferences, recipient evaluation, data release, history, data management
- SIPAS: Transparency, comprehensibility, predictability
- SUS: General Usability
- UEQ: Attractiveness, transparency, efficiency, predictability, stimulation, originality

Qualitative evaluation: Advantages & disadvantages

Quantitative assessment: School grade

Post-measurement of self-efficacy expectation

Procedure 5/5

Investigation

Task 3: App isomorphism
Getting to know the original app (5min)
Getting to know the alternative apps (10min)

Comparison of proposal 1 VAE and UMAP procedure
- Selection judgment and probability assessment for replacement of the original app
- Similarity evaluation by means of rating for both proposals

Refreshing the original app (3min)
Getting to know the alternative apps (10min)

Comparison of proposal 2 VAE and UMAP procedure
- Selection judgment and probability assessment for replacement of the original app
- Similarity evaluation by means of rating for both proposals

Farewell and payment of the expense allowance or certification of the Versuchpersonenstunden.

The experiment lasted on average 74 min (SD = 7.23, Min = 65.82, Max = 98.41).

Data Processing and Evaluation

Quantitative data analysis

- Calculation of descriptive statistics depending on data level (mean = M, standard deviation = SD, minimum = Min, maximum = Max, median = Mdn, absolute frequencies = (X))

- Test for normal distribution

- Selection of non-/parametric methods
  - Differences in behavior levels tested one-sided if directed hypotheses exist otherwise two-sided

Qualitative data analysis

- Classification of the open answers into individual statements

- Creation of a category system (bottom-up, max. two levels: 1. main category, 2. detailed content)

- Allocation of statements to categories by two coders and determination of intercoder reliability

- Decision in the case of deviating coding by the 3rd coder

- Specification of relative frequencies of the answers

- Creation of diagrams with selected sample quotations for frequently used categories

https://iconmonstr.com/file-32-png/

Results

Legend: Männlich / Weiblich / Divers

28,95% — Männlich
68,42% — Weiblich
2,63% — Divers

Age: *M* = 23.95 (*SD* = 5.03; Min = 18.00; Max = 41.00)

Gender : 26 female, 11 male, 1 diverse

Highest level of education : 1.) Gymnasium/Abitur/Fachhochschulreife (27),
2.) University degree (8),
3.) Vocational training (1)

Current employment : 1.) Student (35), 2.) Employee

Courses of study (*n* = 35): 1. Psychology (24), 2. Media Communication (4), 3. Sensor Technology and cog . Psy (4)
4. Public Health (1), Chemistry (1), Human Movement Science (1)

Semester (*n* =35): *M* =2.57 (*SD* =2.33; Min=1; Max=10)

Behavior Levels: VS 1 = 4, VS 2 = 14, VS 3 = 8, VS 4 = 12

Matching and mismatching variant: Matching = 19, mismatching = 19
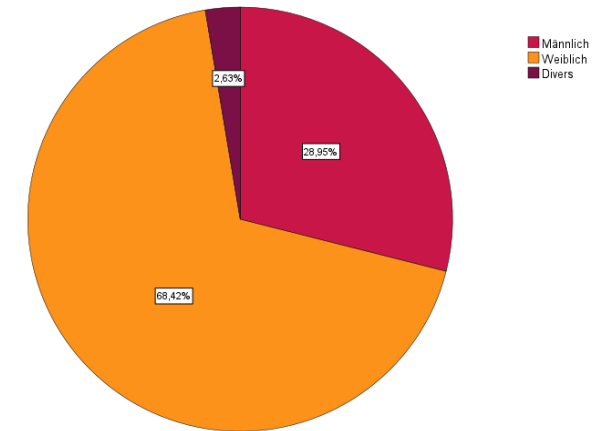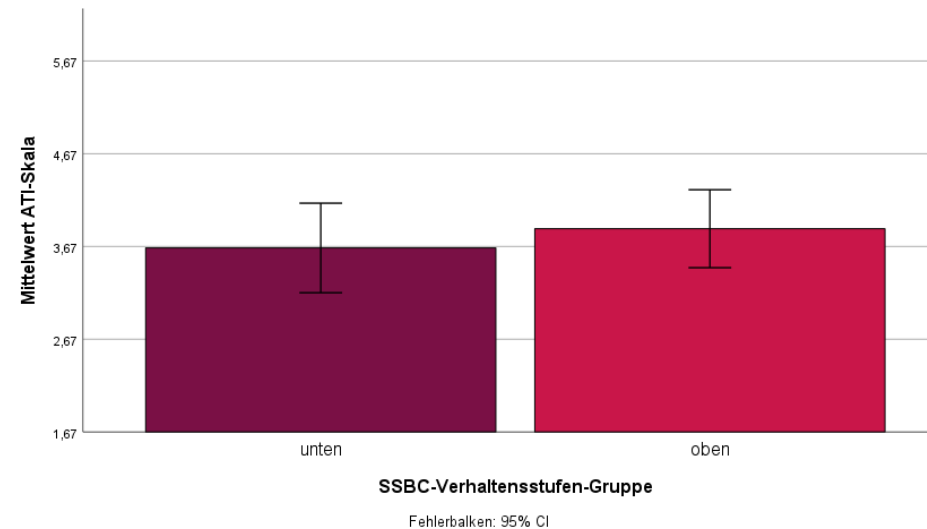
Results: Affinity for technology (ATI Scale; [09])

Average Agreement: M = 3.76 (= "***rather agree***"; SD = .93; Min = 1.67; Max = 5.78)

Comparison with norm sample (N = 300; M = 4.14): significant differences

Our participants consider themselves **less tech-savvy** than a comparable norm sample.

Normal distribution: given

Group differences (two-tailed): **No**

Mittelwert ATI-Skala

SSBC-Verhaltensstufen-Gruppe

Fehlerbalken: 95% CI

[09] Franke, T., Attig, C., & Wessel, D. (2019); https://iconmonstr.com/smartphone-15-png/

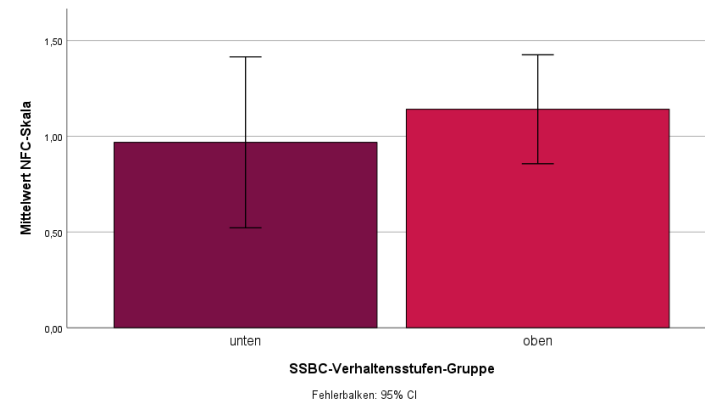# **Results:** Need for cognition [10]

Average agreement: *M* = 1.06 (= "***rather agree***"; *SD* = .75; Min = -1.07; Max = 2.39)

Comparison with the scale mean (0): significantly different

Our participants rate their need for cognition rather high, i.e. they **like to engage in cognitive challenges**.

Normal distribution: given

Group Differences (two-tailed): **No**

[10] Bless , H., Wänke , M., Bohner, G., Fellhauer, RF, & et al. (1994); https://iconmonstr.com/book-24-png /

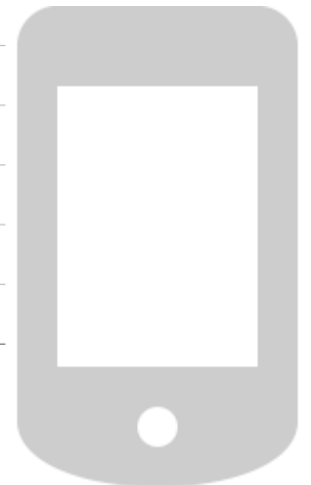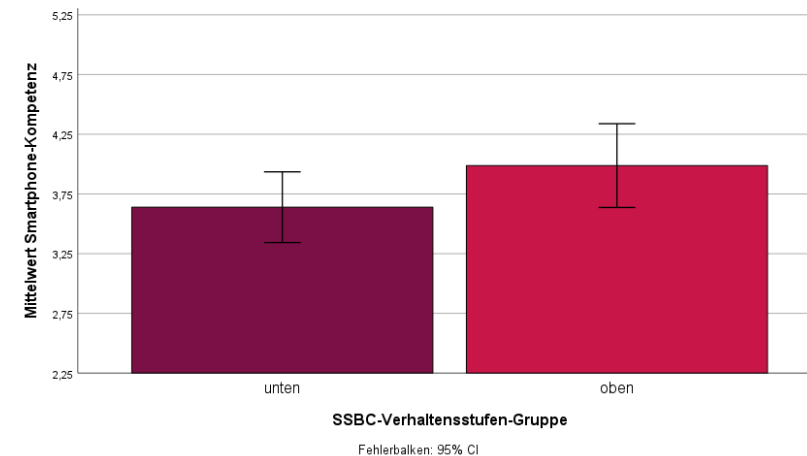# **Results:** Smartphone competence (TAEG; [11])

Average agreement: *M* = 3.82 (= "***rather agree***"; *SD* = .11; Min = 2.25; Max = 5.00).

Comparison with the means of the sample (*N* = 460, *M* = 3.47): significantly different.

Our participants rated their **competence in using smartphones higher** than the norm sample.

Normal distribution: given

Group differences (two-tailed): **No**

Results: App competence (TAEG; [11])

Medium agreement: *M* = 3.64 (= "***rather agree***"; *SD* = .09; Min = 2.25; Max = 4.50)

Comparison with the means of the sample (*N* = 460, *M* = 3.47): significantly different

Our test participants rate their **competence in using apps higher** than the norm sample.

Normal distribution: given

Group differences (two-tailed): **No**

[11] Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009); https://iconmonstr.com/smartphone-7-png/
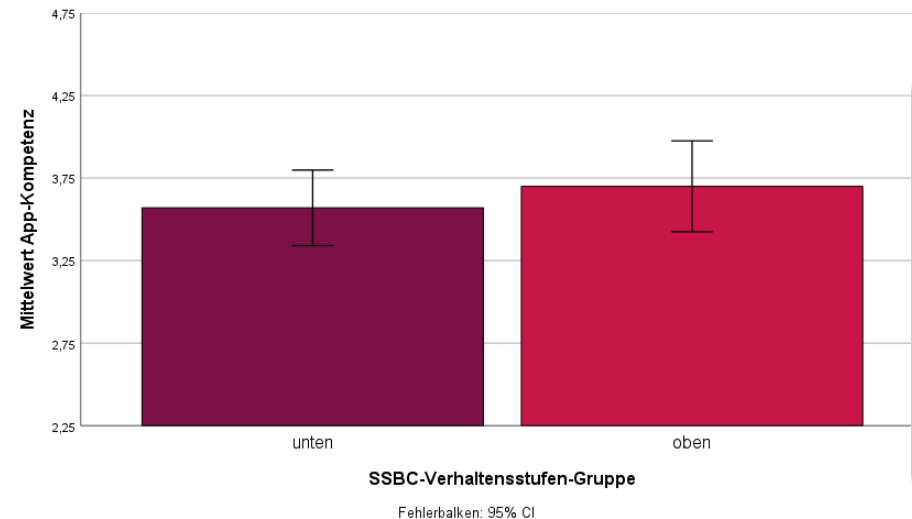
Results: Task 1 performance

Inform about the risk of an app

- Describe the risk factors of this app in your own words (5min)
  - Assignment of points by two raters when mentioning third-party risks, location detection, access to SD card (access to personal data), ID transmission (advertising), security vulnerabilities, forwarding of the data to insecure countries
- Assigned points were added to a score
- In the event of discrepancies between the first coders, a third coder decided
- Cohen's Kappa: first task κ = .75 (= substantial agreement, [12])

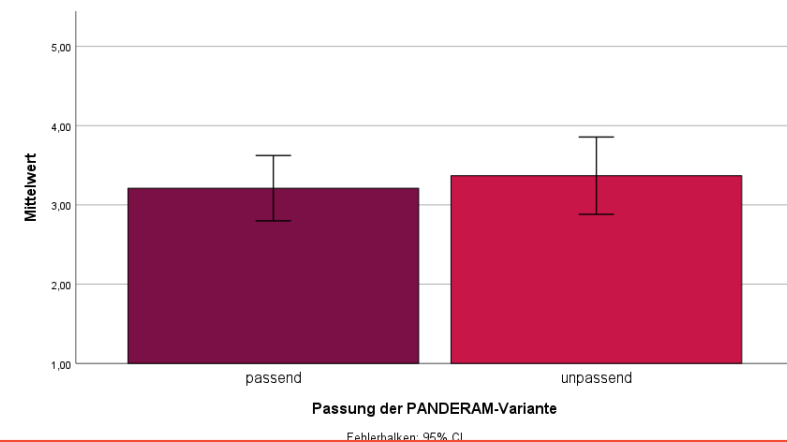# **Results:** Task 1 performance

Inform about the risk of an app

Descriptive Statistics: M = 3.29 (SD = .93, Min = 1.00, Max = 5.00)

Matching condition: n = 19, M = 3.21 (SD = .85, Min = 2.00, Max = 5.00)

Mismatched condition: n = 19, M = 3.37 (SD = 1.01, Min = 1.00, Max = 5.00)

No assumption of normal distribution

One-tailed Mann-Whitney U test : **No difference between the matched and mismatched condition.**

Results: Task 2 performance

Described measures to reduce the risk

- Describe the actions you have taken in your own words (5min).
  - Allocation of points if permissions were revoked, participating apps would be uninstalled or alternative apps installed, background usage was disabled
- Assigned points were added to a score
- If there were discrepancies between the raters, a third rater decided on the final score
- Cohen's Kappa: second task κ = .94 (= almost perfect match, [12])

Results: Task 2 performance

Described measures to reduce the risk

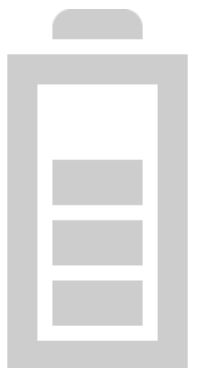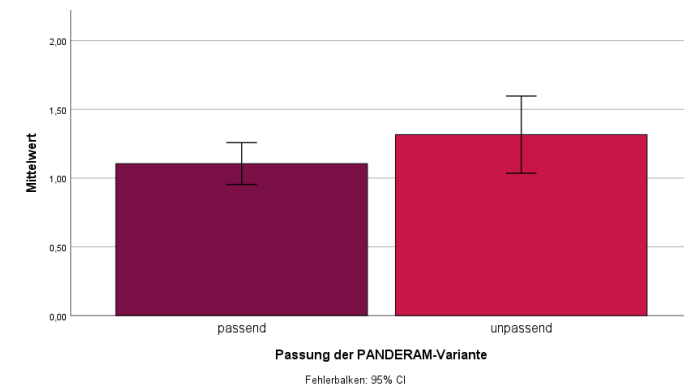Descriptive statistics: M = 1.21 (SD = .47, Min = 1.00, Max = 3.00)

Matching condition: n = 19, M = 1.11 (SD = .32, Min = 1.00, Max = 2.00)

Mismatched condition: n = 19, M = 1.32 (SD = .58, Min = 1.00, Max = 3.00)

No assumption of normal distribution

One-tailed Mann-Whitney U test: **no difference between the matched and mismatched condition.**

https://iconmonstr.com/battery-6-png/

Results: Usability 1/4 (SUS; [13])

Mean SUS Score: *M* =80.00 (*SD* =13.90, Min=30, Max=97.50); corresponds to [14] **grade A-**

Testing against benchmark [14]: SUS score in the lower " *good range* " (score 72.6, lower limit of B+) resulted in a significant difference

Overall, the evaluation of the PANDERAM prototype in terms of **usability is very good** .

Matching condition: *n* = 19, *M* = 4.33, *SD* = .51, *Mdn* = 4.56, *Min* = 2.78, *Max* = 5.00

Mismatched condition: *n* = 19, *M* = 4.11, *SD* = .66, *Mdn* = 4.11, *Min* = 2.11, *Max* = 5.00

No assumption of normal distribution

One-tailed Mann-Whitney U test: **No difference between the matched and mismatched condition.**

[13] Brooke, J. (1996) ; [14] Lewis & Sauro , 2018; https://iconmonstr.com/check-mark-16-png/
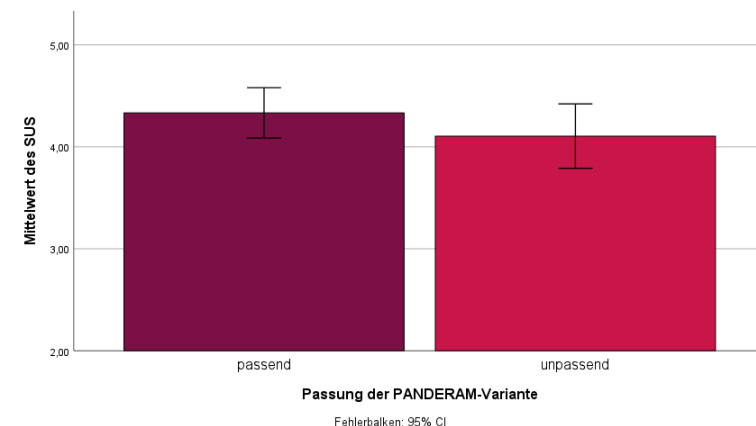
Results: Usability 2/4 (PET-USES; [15])

scale	n	M	SD	Min	Max	mean AndProtect (N =22)	One-tailed t-test vs. mean AndProtect
Privacy Preferences	38	4.25	.46	3.50	5.00	3.64	$t(37) = 8.15, p < .001, d = 1.32$
Recipient Evaluation	38	3.86	.73	1,00	5,00	3.55	$t(37) = 2.60, p = .007, d = .42$
Data Release	38	4.26	.43	3.33	5.00	3.73	$t(37) = 7.72, p < .001, d = 1.25$
History	38	3.47	.45	2.33	4.33		
Data Management	38	3.45	.84	1.50	5.00		
General Usability	38	4.21	.64	2.10	5.00	3.99	$t(37) = 2.12, p = .021, d = .34$

The scales were all answered **positively** ("*agree*") and show significantly **higher values compared to the AndProtect app.**

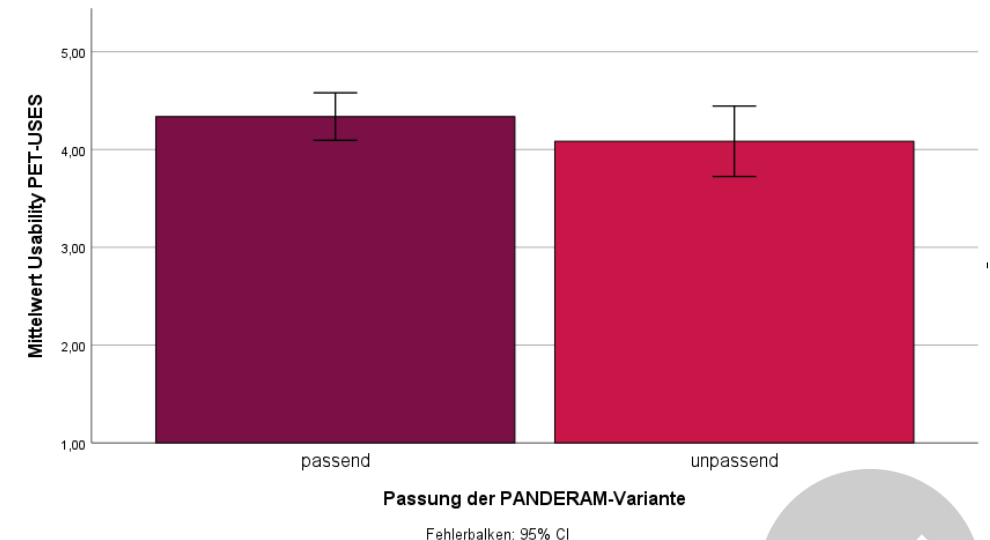[15] Wästlund et al., 2010 ; https://iconmonstr.com/check-mark-16-png/

Results: Usability 3/4 (PET-USES; [15])

Matching condition: n = 19, M = 4.34, SD = .50, Min = 3.10, Max = 5.00

Mismatched condition: n = 19, M = 4.08, SD = .75, Min = 2.10, Max = 5.00

Assumption of the normal distribution

One-tailed t-test for independent samples: **No difference between the matched and mismatched condition.**

Results: Usability 4/4 (meCUE 2.0; [16])

Descriptive statistics for the **Usability subscale**: M = 5.72 (SD = .99, Min = 3.00, Max = 7.00).

One-tailed t-test against scale value 5.00: **Rather positive assessment of usability.**

Matching condition: n = 19, M = 5.60, SD = .95, Mdn = 5.67, Min = 3.00, Max = 7.00

Mismatched condition: n = 19, M = 5.84, SD = 1.03, Mdn = 6.00, Min = 3.33, Max = 7.00

No assumption of normal distribution

One-tailed Mann-Whitney U test: **No difference between the matched and mismatched variant.**

Passung der PANDERAM-Variante

Fehlerbalken: 95% CI

[16] Minge & Riedel, 2013 ; https://iconmonstr.com/check-mark-16-png/
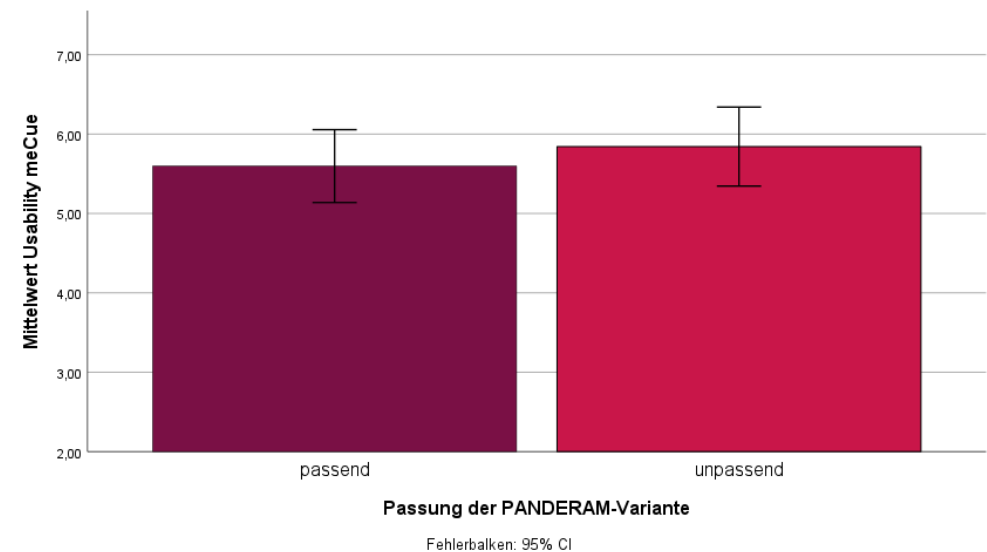
Results: User Experience 1/4 (UEQ; [17])

scale	n	M	SD	Min	Max	Limit benchmark	Unilateral test against lower value evaluation category	Evaluation
Attractiveness	38	1.49	.72	-.83	3	1.17-1.52	$t(37) = 2.74, p = .005, d = 44$	Over the average
Transparency	38	1.88	1.01	-2.25	3	1.56-1.90	$t(37) = 1.92, p = .032, d = 3.1$	Good
Predictability	38	1.80	1.80	.50	2.75	=> 1.65	$t(37) = 1.63, p = .056$	Good
Stimulation	38	1.13	.77	2.50	3.25	.99-1.31	$t(37) = 1.08, p = .143$	Below average
Originality	38	1.07	.89	1.50	2.75	1.05-1.40	$t(37) = .155, p = .439$	Above average
Efficiency	38	1.72	.65	-.25	3	1.47-1.78	$t(37) = 2.41, p = .011, d = .39$	Good

UX is predominantly rated as **"good"** or **"above average"** , only **stimulation** is rated **below average.**

[17] Laugwitz et al., 2008 ; https://iconmonstr.com/smiley-1-svg/

Results: User Experience 2/4 (UEQ; [17])

Matching condition: *n* = 19, *M* = 6.18, *SD* = .56, *Mdn* = 6.00, Min = 5.00, Max = 7.00

Mismatched condition: *n* = 19, *M* = 5.50, *SD* = .93, *Mdn* = 5.50, Min = 3.50, Max = 7.00

No assumption of normal distribution

One-tailed Mann-Whitney U test: A difference **was found between the matched and mismatched condition on the efficiency subscale.**

[17] Laugwitz et al., 2008 ; https://iconmonstr.com/smiley-1-svg/
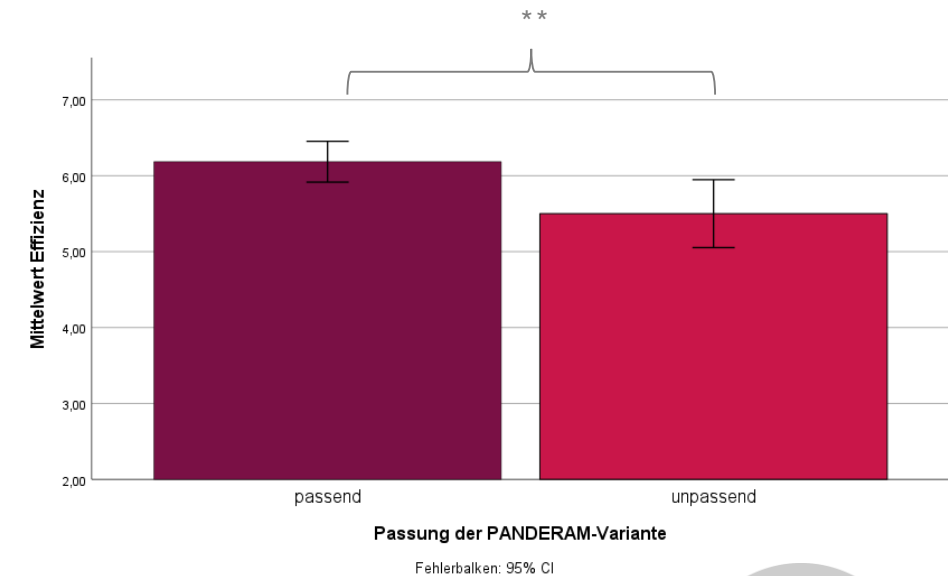
Results: User Experience 3/4 (meCUE 2.0; [16])

Descriptive statistics for the **positive emotions subscale**: M = 3.94 (SD = .91, Min = 1.5, Max = 5.33).

One-tailed t-test against scale value 4.00: not significant.

Neutral assessment on the "positive emotions" subscale (neither/nor).

Descriptive statistics for the **negative emotions subscale**: M = 2.37 (SD = .1.03, Min = 1, Max = 5.83).

One-tailed t-test against scale value 3.00: significant difference.

Rejection for the subscale "negative emotions".

Participants do not experience any negative emotions when using the app.
The experience of positive emotions is described as neutral.

Results: User Experience 4/4 (meCUE 2.0; [16])

Positive emotions

Matching condition: n = 19, M = 4.10, SD = .75, Min = 2.50, Max = 5.33

Mismatched condition: n = 19, M = 3.79, SD = 1.05, Min = 1.50, Max = 5.33

Assumption of the normal distribution

One-tailed t-test for independent samples: **No difference found between the matched and mismatched variant.**

Negative emotions

Matching condition: n = 19, M = 2.12, SD = .76, Min = 1.00, Max = 3.50

Mismatched condition: n = 19, M = 2.62, SD = 1.21, Min = 1.00, Max = 5.83

Assumption of the normal distribution

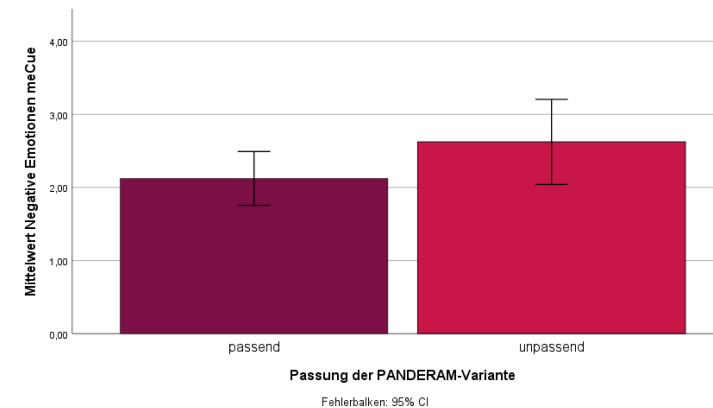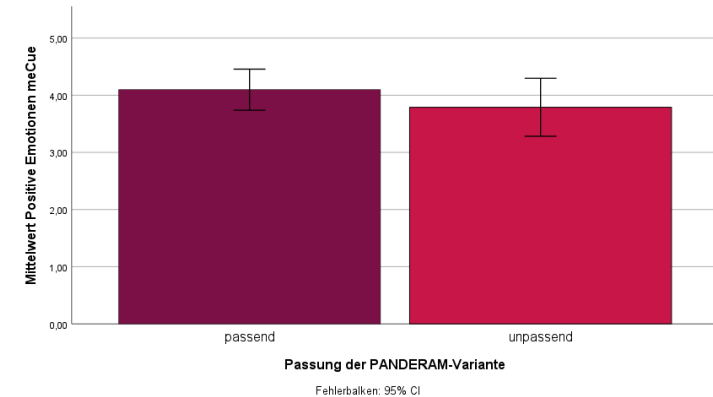One-tailed t-test for independent samples: **No difference found between the matched and mismatched variants.**

[16] Minge & Riedel, 2013 ; https://iconmonstr.com/smiley-1-svg/

Results: Transparency (SIPAS; [18])

Compared to the scale mean (3.5), there is a significant difference for predictability (small effect) and significant differences for transparency and comprehensibility (large effects).

SIPAS - subscale	T value	df	Significance (one-sided)	Effect size Cohen d
transparency	9.50	37	<.001	1.54
intelligibility	6.12	37	<.001	.99
predictability	1.92	37	.031	.31

The **PANDERAM app** is rated above average regarding **transparency, understandability and predictability.** .

[18] Schrills & Franke, 2021; https://iconmonstr.com/smiley-1-svg/

Results: Self-efficacy expectation 1/2 [19]

Measurement	n	M	SD	MIN	MAX
pre-measurement	38	2.38	.54	1.22	3.44
post-measurement	38	2.52	.50	1.30	3.70

T-test (one-tailed) before and after measurement: significant difference.

After presentation and interaction with the PANDERAM app, participants report **higher self-efficacy expectations**, i.e., they feel more empowered to act on data protection and privacy problems afterwards.

[19] Schwarzer, R., & Jerusalem, M. (2002) ; https://iconmonstr.com/wrench-24-png/
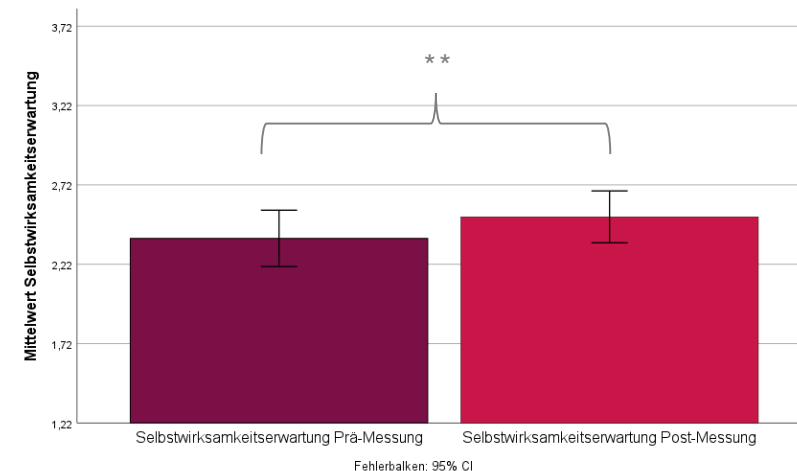
Results: Self-efficacy expectation 2/2 [20]

Measurement	n	M	SD	MIN	MAX
pre-measurement	38	6.13	2.15	1	10
post-measurement	38	6.45	2.01	3	10

T-test (one-tailed) before and after measurement: significant difference.

Participants also report **higher self-efficacy expectation regarding data protection and privacy problems** after the presentation and interaction with the PANDERAM app.
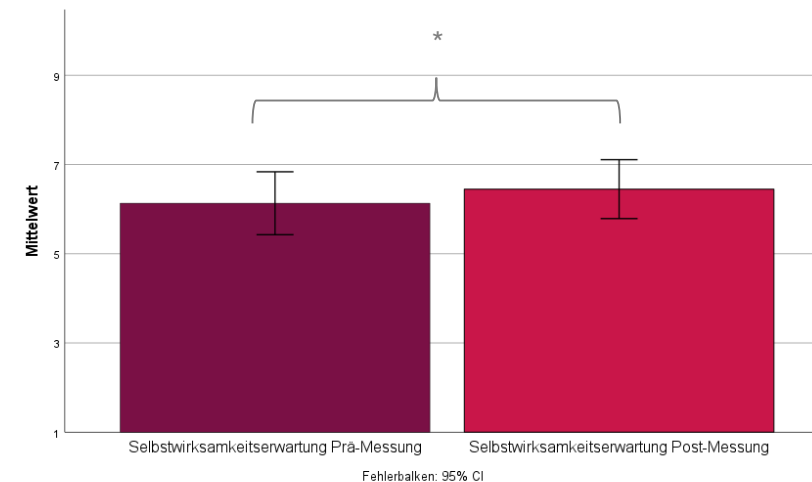
Results: Intention to use (UTAUT; [21])

Measurement	n	mean	SD	MIN	MAX	Unilateral testing against the mean of the scale
Performance Expectation	38	3.20	1.12	1.00	5.00	$t\,(37) = -4.38, p <.001, d = -.71$
Effort Expectation	38	5.58	1.25	1.00	7.00	$t\,(37) = 7.81, p < .001, d = 1.27$
Attitude towards technology	38	4.72	.81	1.75	6.25	$t\,(37) = 5.49, p <.001, d = .89$
Social influence	38	4.91	.93	3.00	6.75	$t\,(37) = 6.07, p < .001, d = .99$
Supporting Conditions	38	5.18	.98	1.67	6.67	$t\,(37) = 7.38, p < .001, d = 1.20$
Self-efficacy expectation	38	4.95	1.24	1.25	7.00	$t\,(37) = 4.75, p <.001, d = .77$
Anxiety	38	2.39	1.14	1.00	5.75	$t\,(37) = -8.70, p <.001, d = -1.41$
Intention of use	38	5.09	1.11	1.00	7.00	$t\,(37) = 6.03, p < .001, d = .98$

The **expected performance** of the app is **negative**. The **other aspects** of the app (e.g. expectation of how strenuous the app is or anxiety) are **rated positively.**

[21] Venkatesh, V., Morris, MG, Davis, GB, & Davis, FD (2003); https://iconmonstr.com/smiley-1-svg/

Results: Praise for the app (qualitative)

Which aspects of the PANDERAM app did you **particularly like**?
Name the three most important.

121 statements were coded in categories with two parallel levels: 1st level = distinction between "information" (37% of the statements), "design" (35%), and "options for action" (28%).

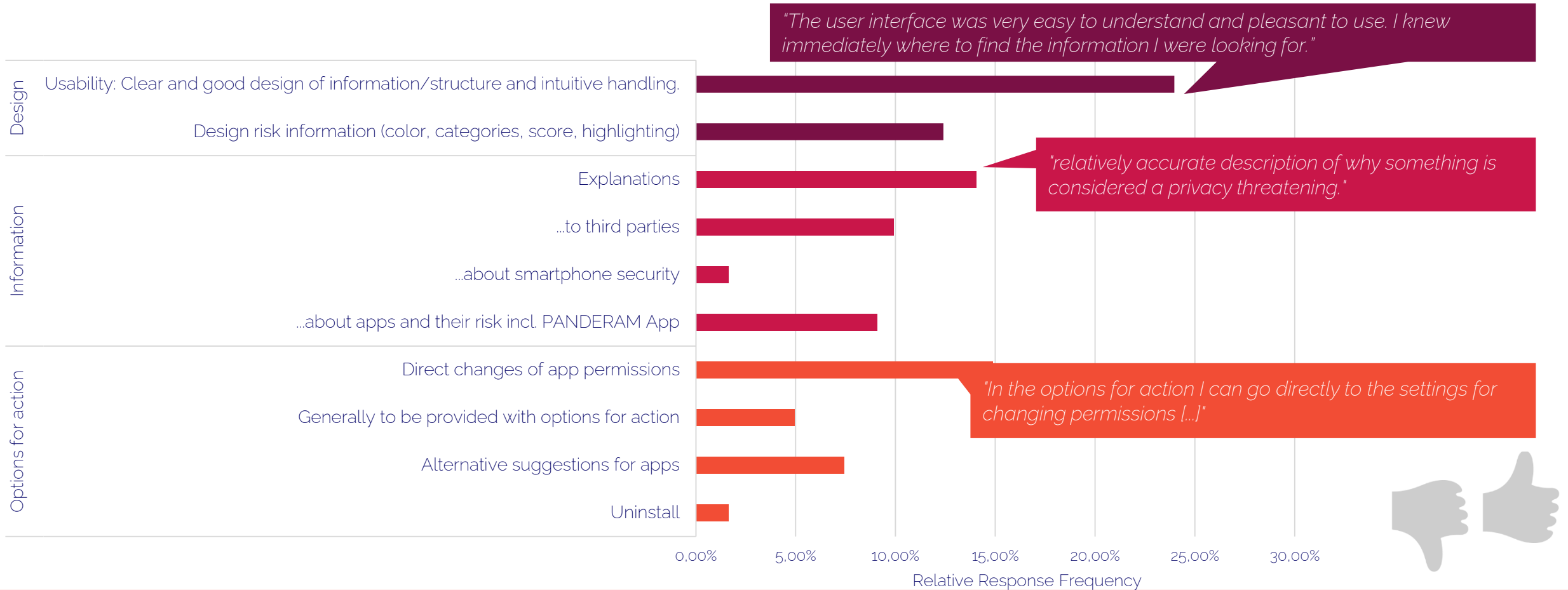2nd level: More detailed content (see next slide)

Intercoder reliability (weighted):
κ = 0.84 (= (almost) perfect match in coding; [12])

[12] Landis, JR, & Koch, GG (1977) ; https://iconmonstr.com/thumb-5-png/; https://iconmonstr.com/thumb-1-png/

Results: Praise for the app (qualitative)

"The user interface was very easy to understand and pleasant to use. I knew immediately where to find the information I were looking for."

Design

- Usability: Clear and good design of information/structure and intuitive handling.
- Design risk information (color, categories, score, highlighting)

"relatively accurate description of why something is considered a privacy threatening."

Information

- Explanations
- ...to third parties
- ...about smartphone security
- ...about apps and their risk incl. PANDERAM App

"In the options for action I can go directly to the settings for changing permissions [...]"

Options for action

- Direct changes of app permissions
- Generally to be provided with options for action
- Alternative suggestions for apps
- Uninstall

Axis: 0,00% 5,00% 10,00% 15,00% 20,00% 25,00% 30,00%

Relative Response Frequency

34

Results: Criticism for the app (qualitative)

What **opportunities for improvement** do you see for the PANDERAM app?
Name the three most important things.

93 statements were coded in categories with two parallel levels: 1st level = differentiation between "**extend the scope of information and functions of the app**" (44% of the statements), "**improve the design of existing information**" (43%). 9% of the answers accounted for "none" suggestions for improvement.
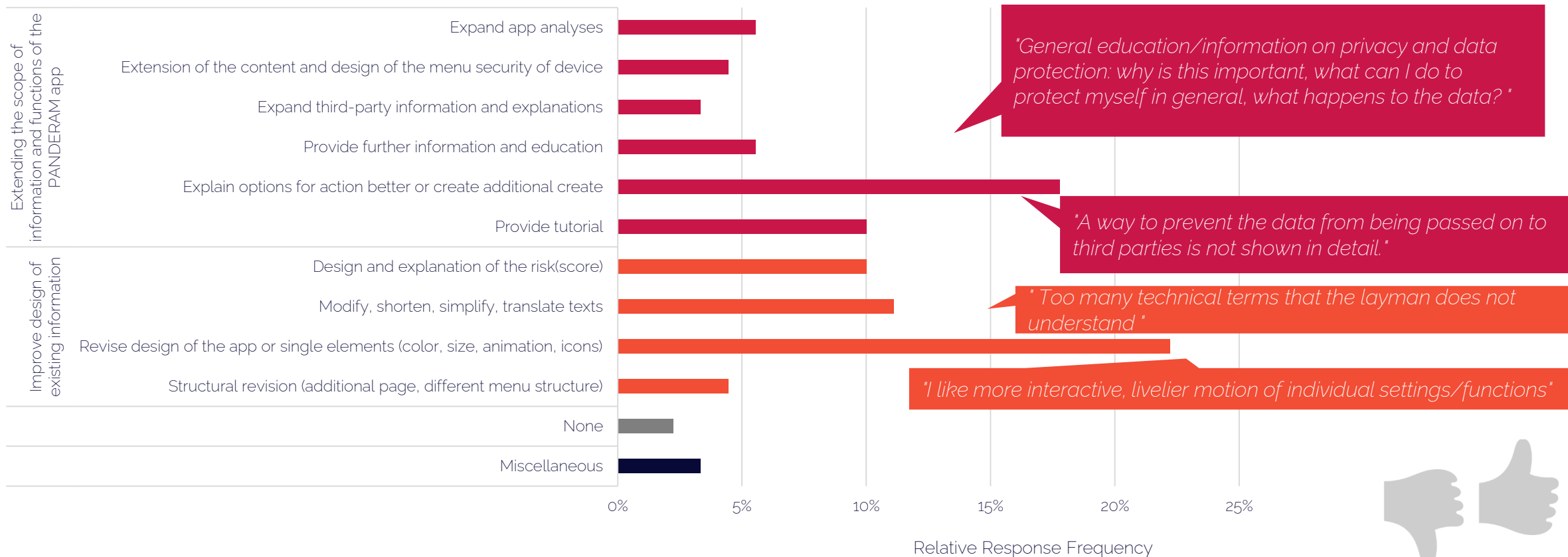
2nd level: More detailed content (see next slide)

Intercoder reliability (weighted):
κ = 0.65 (= substantial agreement; [12])

[12] Landis, JR, & Koch, GG (1977) ; https://iconmonstr.com/thumb-5-png/; https://iconmonstr.com/thumb-1-png/

Results: Criticism for the app (qualitative)

"General education/information on privacy and data protection: why is this important, what can I do to protect myself in general, what happens to the data?"

"A way to prevent the data from being passed on to third parties is not shown in detail."

"Too many technical terms that the layman does not understand"

"I like more interactive, livelier motion of individual settings/functions"

Extending the scope of information and functions of the PANDERAM app
- Expand app analyses
- Extension of the content and design of the menu security of device
- Expand third-party information and explanations
- Provide further information and education
- Explain options for action better or create additional create
- Provide tutorial

Improve design of existing information
- Design and explanation of the risk(score)
- Modify, shorten, simplify, translate texts
- Revise design of the app or single elements (color, size, animation, icons)
- Structural revision (additional page, different menu structure)
- None
- Miscellaneous

Relative Response Frequency

0% 5% 10% 15% 20% 25%

Results: School grade

What grade would you give the PANDERAM app overall?

On a scale of (1 = *very good* , 2 = *good* , 3 = *satisfactory* , 4 = *sufficient* , 5 = *poor* and
6 = *insufficient*), the *N* = 38 participants in the laboratory experiment gave the mean grade "***good*** "
(*Mdn* = 2.00; Min = 1.00; Max = 5.00).

The two **behavioral level groups do not differ** in their rating. The experimental condition
also makes **no difference to the grading**.

Results: Comparison of app isomorphism methods

Exit app: AirBrush

Ranking/ procedure	UAE		UMAP
1st place	PicShot	⟷	Ultimate photo blender
place 2	Photo Editor Pro	⟷	Photo Blender

38

Results: Comparison of app isomorphism methods

Which of the two suggestions do you feel is more similar to the original app?

The vast majority (86.8%; 33) of the participants judged that
PicShot (VAE1) is more similar to AirBrush than Ultimate Photo Mixer (UMAP 1).

Depending on the previous selection, the participants should answer the following:

1.) If [PicShot/Ultimate Photo Mixer] is less risky for my privacy, I would replace the original AirBrush app with this alternative app. (1 = *completely disagree*, 2 = *disagree*, 3 = *agree*, 4 = *completely agree*)

- Answers for PicShot (UAE 1, n =33): M = 2.58 (SD = 0.90, Min = 1.00, Max = 4.00)

- Answers for Ultimate Photo Mixer (UMAP 1, n = 5): M = 2.00 (SD = 0.00, Min = 2.00, Max = 2.00)

Since the groups are very unequally occupied, no comparisons can be calculated here.

Results: Comparison of app isomorphism methods

In addition, the participants were asked to give a similarity rating for both apps separately (from 1 = *not at all similar* to 10 = *very similar*).
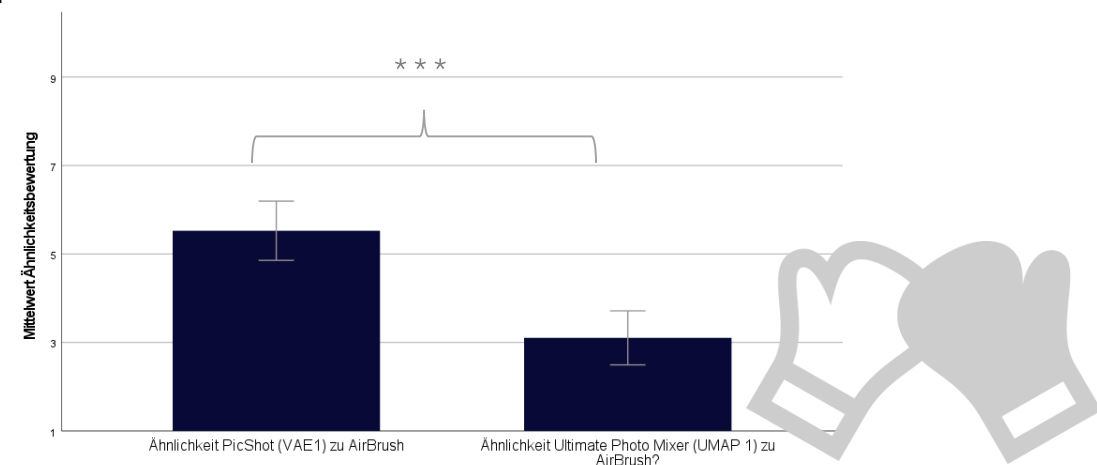
Similarity PicShot (VAE 1) to AirBrush: *M* = 5.53 (*SD* = 2.04, Min = 1.00 , Max = 9.00)

Similarity Ultimate Photo Mixer (UMAP 1) to AirBrush: *M* = 3.11 (*SD* = 1.86, Min = 1.00, Max = 7.00)

These assessments are not normally distributed in each case

The similarity ratings of **PicShot and Ultimate Photo Mixer differ significantly** . **PicShot** (*Mdn* = 6.00) **is rated significantly more similar to AirBrush** (*Mdn* = 2.50) than Ultimate Photo Mixer.

40

Results: Comparison of app isomorphism methods

Which of the two suggestions do you feel is more similar to the original app?

The vast majority (84.2%; 32) of the participants judged that
Photo Editor Pro (VAE2) is more similar to AirBrush than Photo Blender.

Depending on the previous selection, the participants should answer the following again:

Unless [Photo Editor Pro/Photo Blender] is less risky for my privacy, I would replace the original AirBrush app with this alternative app. (1 = *completely disagree*, 2 = *disagree*, 3 = *agree*, 4 = *completely agree*)

* Anwerss for Photo Editor Pro (UAE 2, *n* = 32): *M* = 2.81 (*SD* = 0.90, Min = 1.00, Max = 4.00)

* Answers for Photo Blender (UMAP 2, *n* = 6): *M* = 3.17 (*SD* = 1.33, Min = 1.00, Max = 4.00)

Since the groups are very unequally occupied, no comparisons can be calculated here.

Results: Comparison of app isomorphism methods

In addition, the participants were asked to give a similarity rating for both apps separately
(1 = *not at all similar* - 10 = *very similar*).

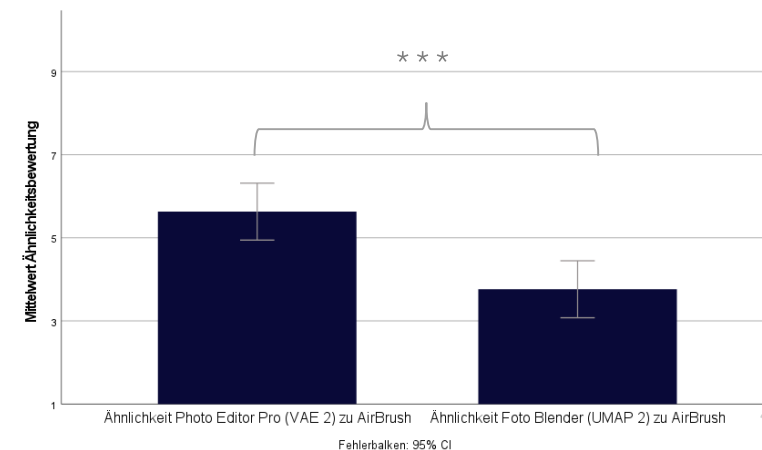Similarity Photo Editor Pro (UAE 2) to AirBrush: *M* = 5.63 (*SD* = 2.09, Min = 2.00, Max = 9.00)

Similarity Photo Blender (UMAP 2) to AirBrush: *M* = 3.76 (*SD* = 2.09, Min = 1.00, Max = 9.00)

These assessments are not normally distributed in each case.

The similarity ratings of **Photo Editor Pro and
Photo Blender differ statistically significantly.
Photo Editor Pro** (*Mdn* = 6.00) is judged
to be much more similar to AirBrush than
Photo Blender (*Mdn* = 3.00).

Mittelwert Ähnlichkeitsbewertung

Ähnlichkeit Photo Editor Pro (VAE 2) zu AirBrush   Ähnlichkeit Foto Blender (UMAP 2) zu AirBrush

Fehlerbalken: 95% CI

Results: Comparison of app isomorphism methods

Comparison of similarity ratings UAE

The similarity ratings of **PicShot** (VAE 1, *Mdn* = 6.00) **and Photo Editor Pro** (VAE 2, *Mdn* = 6.00) **do not differ** statistically.

Comparison similarity ratings UMAP

The similarity ratings of **Ultimate Photo Mixer** (UMAP 1, *Mdn* = 2.50) **and Photo Blender** (UMAP 2, *Mdn* = 3.00) **differ** statistically. The **second suggestion** is **evaluated as more similar than the first suggestion**, which the method delivers.

https://iconmonstr.com/boxing-2-png/; https://iconmonstr.com/boxing-1-png/;

43
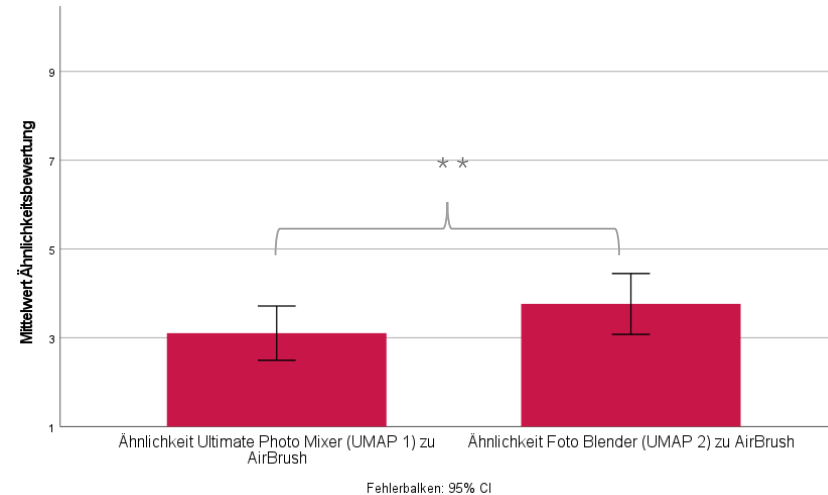
Summary

App rating

- The **usability** of the PANDERAM app is rated as positive or "**very good**".

- The **user experience** as "above average" or "**good**"

  - However, users rate the stimulation of the app as a weakness.

  - The experience of positive emotions is classified as neutral, while the experience of negative emotions is rejected.

- The **transparency** of the PANDERAM app is rated as **above average**

- The **PANDERAM app increases the self-efficacy expectation** of the participants to be able to deal with data protection problems.

- The **intention to use the app is positive** overall, but participants do not assume that the app can improve their (work) performance

Summary

App rating

- The participants most frequently demand for:

  - **Better explained and more options for action**, for example to prevent **the data from being passed on to third parties**

  - More general **information and education on the subject privacy protection** and a **tutorial** for first use

  - In addition, a revision of individual elements, such as a **more interactive design** of individual settings/functions, is desired

  - The texts used were criticized, for example with regard to **technical terms**

- Most frequent praise: the user interface was rated as **user-friendly**, **explanations** and **direct action option** were also praised

- The app was rated with the **school grade "good"**

Summary

Comparison of matching vs. mismatching app variant

- Whether the participants were presented with a matching or mismatching prototype for their behavioral level was irrelevant for usability and performance

- In the user experience subscales, only the "**efficiency**" subscale differs between participants, who receive a matching or mismatching variant

  - That is, the app was perceived as more inefficient, unpragmatic, or cluttered for participants in the mismatching condition than for participants in the matching condition

Summary

Laboratory experiment app isomorphism

- In both judgments that the participants were asked to make, they chose the proposed **app by the VAE method** and judged it to be **more similar to the original app**

- For the two similarity ratings, participants rated the **VAE method proposal significantly more similar than the UMAP** proposal

- Comparing the 1st and 2nd proposal of each method, there were no differences for VAE; for UMAP, the 2nd proposal provided by the method was rated as slightly more similar than the 1st proposal.

From the perspective of the participants, the VAE procedure

provides better suggestions of similar apps.

Outlook

- Based on the results, the PANDERAM app will be revised
- The new version is tested again in a smaller laboratory study
- Here, the focus will be on IT-security issues of the app.

Thank you for your attention!

Susen Döbelt

Wilhelm-Raabe-Str. 43
09120 Chemnitz

Telephone: 0371 531 33615
Email: susen.doebelt@psychologie.tu-chemnitz.de