



PANDERAM

Laborstudie 1 PANDERAM Prototyp App-Analyse und -Isomorphie Ergebnisse AP 7.2 und 7.3

Susen Döbelt & Dominik Lange

Chemnitz, April 2023



Überblick



Einleitung: Ziele, Organisation, Ablauf, Datenaufbereitung und Auswertung
Ergebnisse

- **Laborstudie Aufgabe 1 und 2**
- **Laborstudie Aufgabe 3: App-Isomorphie**

Zusammenfassung

Ausblick



Ziel(e) der Befragung

Untersuchung des PANDERAM-Prototyps hinsichtlich Verbesserungspotential für **Usability** und **User Experience** sowie **Passung** und **Wirksamkeit** für unterschiedliche **Verhaltensstufen**.

Forschungsfragen:

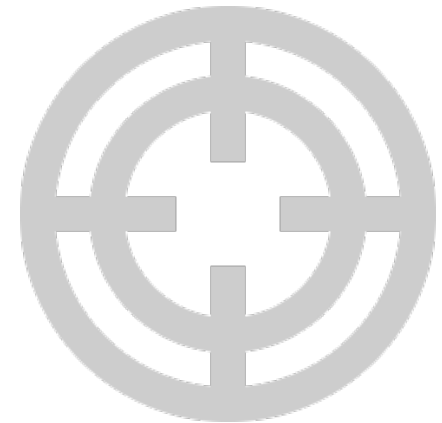
Wie bewerten Personen unterschiedlicher Verhaltensstufen den zur Verhaltensstufe passenden vs. nicht passenden Prototyp?

Steigert der PANDERAM Prototyp die Selbstwirksamkeitserwartung in Bezug auf Daten- und Privatsphärenschutz?

Wie bewerten Nutzende die Transparenz des PANDERAM Prototypen?

Können von der Verhaltensstufe, App-Variante und der Transparenz Rückschlüsse auf die Nutzungsintention gezogen werden?

Welches App-Isomorphie-Verfahren liefert aus Nutzersicht die besseren App-Alternativ-Vorschläge?

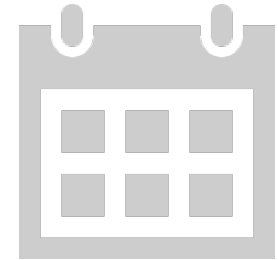




Organisation 1/2

Zeitlicher Ablauf

- Beginn Konzeption (05/2022)
- Implementierung und Testung Fragebögen (07/2022)
- Start (Rekrutierung): 13.11.2022
- Start (Laborversuch): 07.12.2022
- Abschluss: 31.01.2023 (5 Wochen Versuchsdurchführung)





Organisation 2/2

Rekrutierung

- Studienaufruf mit Link zum Rekrutierungsfragebogen
- **Vergütung:** Ausfüllen Rekrutierung: 0,5 Versuchspersonenstunden bzw. Teilnahme an Verlosung von 5 x 20€, Vergütung für den Laborversuch: 1,5 Versuchspersonenstunden oder 25€ in bar)
- **Verbreitung:** über Studienteilnahmeverteiler TUC (mehrere Erinnerungen), AHF Webseite, Aushänge in der Wilhelm-Raabe-Str./TU Campus/Mensa/privates Umfeld, Chat- und Facebook Gruppen privates Umfeld



Den Rekrutierungsfragebogen füllten insgesamt $N= 129$ Personen vollständig aus.

- Kriterien für die Auswahl Laborversuch:
 - Android Betriebssystem >6.0 ($n = 70$, = 54% aus dem Rekrutierungsfragebogen)
 - Verhaltensstufenzuordnung möglich ($n = 69$) und Ausgewogenheit der Verhaltenstufengruppen

Letztendlich nahmen $N= 38$ Personen am Laborversuch teil.



Ablauf 1/5

Vor Untersuchungsbeginn

- **Absage** an nicht geeignete Personen aus dem Rekrutierungsfragebogen und **Vergütung** der Angemeldeten
- Zusage via E-Mail an ausgewählte Personen
 - Bitte um **Terminauswahl**
- **Zuordnung zur Versuchsbedingung** (zur Verhaltensstufe passender bzw. unpassender Prototyp) und Fragebogenvariante A oder B (randomisierte Anzeige von Inhalten)





Ablauf 2/5

Untersuchung

- **Willkommen** und Bitte um Lesen und Zeichnen der **Teilnahme- und Datenschutzinformation**
- Information zur Verfügung stehenden **Gerätschaften**: Laptop und Maus zum Ausfüllen des Fragebogens, Test-Smartphone sowie darauf installierter PANDERAM-Prototypen
- **Hinweise** vom Versuchsleiter:
 - PANDERAM App die bewertet werden soll, ist noch ein Prototyp (einige Bereiche sind nicht funktional bzw. noch nicht vollständig ausgearbeitet)
 - Bitte keine Apps vom Test-Smartphone löschen
 - Anweisungen im Fragebogen befolgen und den Timer im Auge behalten
 - Bei aufkommenden Fragen an VersuchsleiterIn wenden

Start des Versuches

- Erfassung selbstbeschreibender Variablen: Technikaffinität, Smartphone-Kompetenz und Need for Cognition
- Prämessung Selbstwirksamkeitserwartung (Zustimmung zu Aussagen und Rating)
- Kennenlernen des PANDERAM Prototypen (10min)



Ablauf 3/5

Untersuchung

Aufgabe 1:

Informieren Sie sich mithilfe des PANDERAM-Prototypen über die App „Freeletics“

- Beschreiben Sie die Risikofaktoren dieser App in eigenen Worten (5min)
- Leistungsmessung Prä: Wie hoch ist der aktuelle Risikoscore der App „Freeletics“?

Aufgabe 2:

Ergreifen Sie mithilfe des PANDERAM-Prototypen Maßnahmen, um das **Risiko** der App „Freeletics“ zu **minimieren**

- Beschreiben Sie die von Ihnen ergriffenen Maßnahmen mit eigenen Worten (5min)
- Leistungsmessung Post: Wie hoch ist der aktuelle Risikoscore der App „Freeletics“ jetzt?





Ablauf 4/5

Untersuchung

Bewertung PANDERAM-Prototyp

Quantitative Bewertung durch:

- UTAUT Leistungserwartung, Anstrengungserwartung, Einstellung gegenüber Technologie
- meCUE 2.0 Usability, Nützlichkeit, Nutzungsintention, positive u. negative Emotionen
- PET-USES General Usability, Privacy Preferences, Recipient Evaluation, Data Release, History, Data Management
- SIPAS Transparenz, Verständlichkeit, Vorhersagbarkeit
- SUS Allgemeine Usability
- UEQ Attraktivität, Durchschaubarkeit, Effizienz, Vorhersagbarkeit, Stimulation, Originalität

Qualitative Bewertung durch Vorteile & Nachteile

Quantitative Bewertung durch Schulnote

Post-Messung der Selbstwirksamkeitserwartung





Ablauf 5/5

Untersuchung

Aufgabe 3: App-Isomorphie

Kennenlernen Original-App (5min)

Kennenlernen der Alternativ-Apps (10min)

Vergleich Vorschlag 1 VAE und UMAP Verfahren

- Auswahlurteil und Wahrscheinlichkeitseinschätzung für Ersatz der Original-App
- Ähnlichkeitsbewertung mittels Rating für beide Vorschläge

Auffrischen der Original-App (3min)

Kennenlernen der Alternativ-Apps (10min)

Vergleich Vorschlag 2 VAE und UMAP Verfahren

- Auswahlurteil und Wahrscheinlichkeitseinschätzung für Ersatz der Original-App
- Ähnlichkeitsbewertung mittels Rating für beide Vorschläge

Verabschiedung und Auszahlung der Aufwandsentschädigung bzw. Bescheinigung der Versuchspersonenstunden

Der Versuch dauerte im Mittel 74 min ($SD = 7.23$, $Min = 65.82$, $Max = 98.41$).





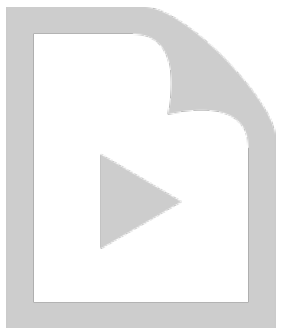
Datenaufbereitung und Auswertung

Quantitative Datenanalyse

- Berechnung deskriptiver Statistiken je nach Datenniveau (Mittelwert = *MW*, Standardabweichung = *SD*, 0 Minimum = Min, Maximum = Max, Median = *Mdn*, absolute Häufigkeiten = (X))
- Testung auf Normalverteilung
- Auswahl non-/parametrischer Verfahren
 - Unterschiede bzgl. Verhaltensstufen einseitig getestet, wenn gerichtete Hypothesen bestehen ansonsten zweiseitig

Qualitative Datenanalyse

- Einteilung der offenen Antworten in einzelne Aussagen
- Erstellung eines Kategoriensystems (Bottom-up, max. zwei Ebenen: 1. Oberkategorie, 2. Detaillierte Inhalte)
- Zuordnung Aussagen zu Kategorien durch zwei KodiererInnen und Bestimmung der Intercoder-Reliabilität
- Entscheidung bei abweichenden Kodierungen durch 3. KodiererIn
- Angabe relativer Häufigkeiten der Antworten
- Erstellung von Diagrammen mit ausgewählten Beispielzitate für häufig besetzte Kategorien





Ergebnisse: Demografie

Alter: $MW = 23.95$ ($SD = 5.03$; Min = 18.00; Max = 41.00)

Geschlecht: 26 weiblich, 11 männlich, 1 divers

Höchster Bildungsabschluss: 1.) Gymnasium/Abitur/Fachhochschulreife (27),
2.) Hochschulabschluss (8),
3.) Berufsausbildung (1)

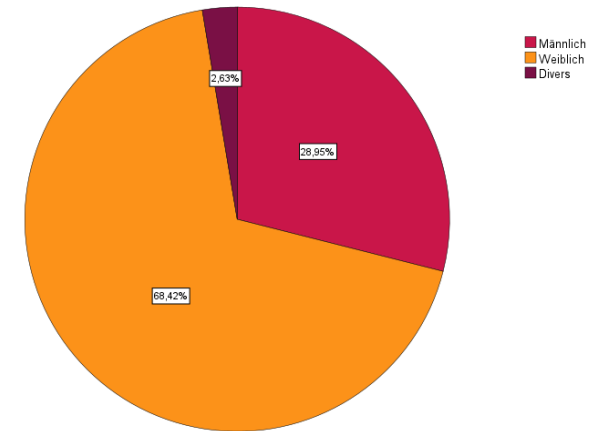
Aktuelles Beschäftigungsverhältnis: 1.) StudentIn (35), 2.) ArbeitnehmerIn

Studiengänge ($n = 35$): 1. Psychologie (24), 2. Medienkommunikation (4), 3. Sensorik und kog. Psy. (4)
4. Public Health (1), Chemie (1), Human Movement Science (1)

Semester ($n = 35$): $MW = 2.57$ ($SD = 2.33$; Min = 1; Max = 10)

Verhaltensstufen: VS 1 = 4, VS 2 = 14, VS 3 = 8, VS 4 = 12

Passende und unpassende Variante: Passend = 19, Unpassend = 19





Ergebnisse: Technikaffinität (ATI Scale; [og])

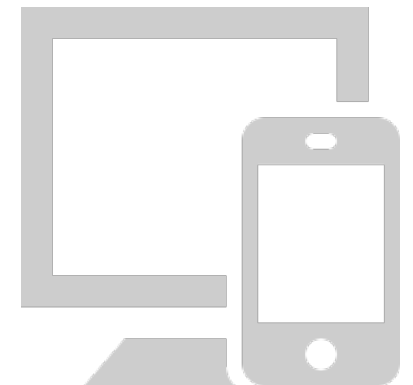
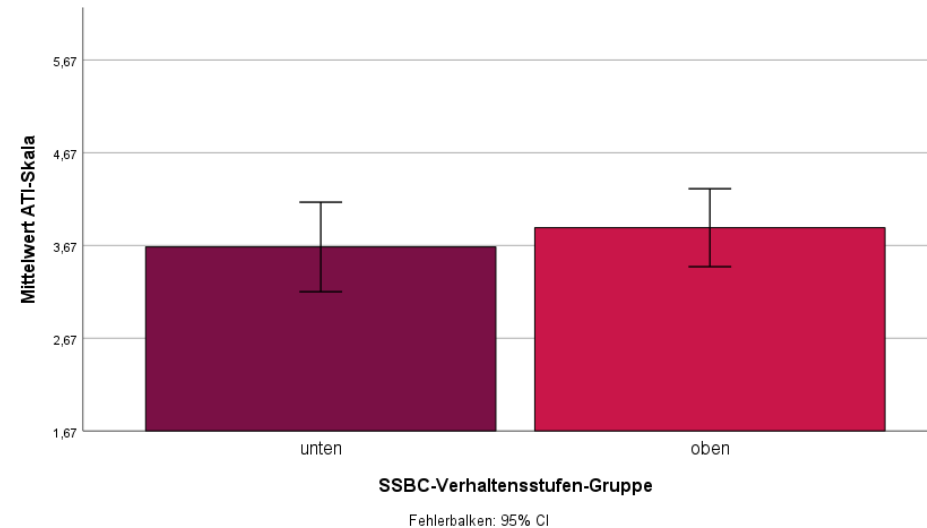
Mittlere Zustimmung: $MW = 3.76$ („**Stimmt eher**“; $SD = .93$; Min = 1.67; Max = 5.78)

Vergleich mit Normstichprobe ($N = 300$; $MW = 4.14$): signifikante Unterschiede

Unsere Versuchspersonen schätzen sich **weniger technikaffin** als eine vergleichbare Normstichprobe ein.

Normalverteilung: gegeben

Gruppenunterschiede (zweiseitig): **Nein**





Ergebnisse: Need for Cognition [10]

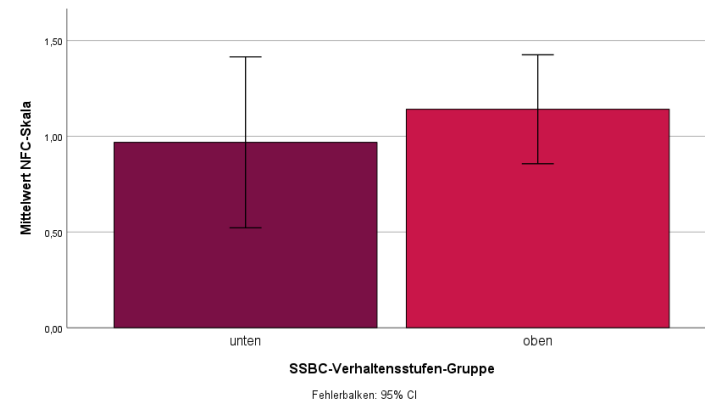
Mittlere Zustimmung: $MW = 1.06$ (= „**trifft eher zu**“; $SD = .75$; Min = -1.07; Max = 2.39)

Vergleich mit den Skalenmittelwert (0): signifikant verschieden

Unsere Versuchspersonen schätzen ihr Kognitionsbedürfnis eher hoch ein, d.h. sie **beschäftigen sich gern mit kognitiven Herausforderungen**.

Normalverteilung: gegeben

Gruppenunterschiede (zweiseitig): **Nein**





Ergebnisse: Smartphone-Kompetenz (TAEG; [11])

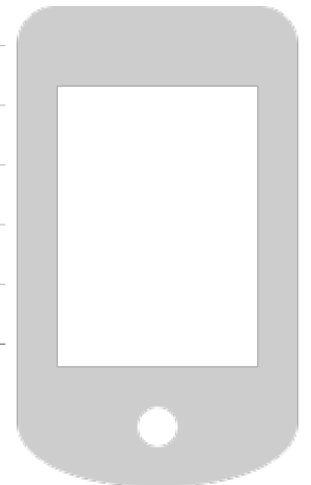
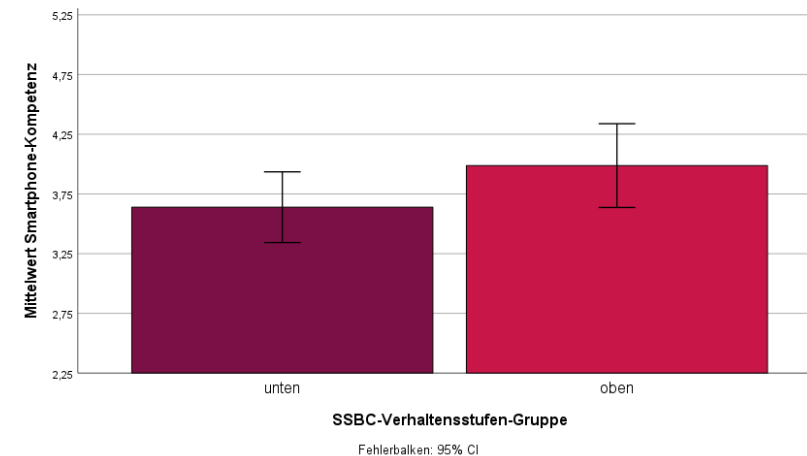
Mittlere Zustimmung: $MW = 3.82$ (= „**trifft eher zu**“; $SD = .11$; Min = 2.25; Max = 5.00)

Vergleich mit den Mittelwerten der Stichprobe ($N = 460$, $MW = 3.47$): signifikant verschieden

Unsere Versuchspersonen schätzen hinsichtlich ihrer **Kompetenz im Umgang mit Smartphones höher** als die Normstichprobe.

Normalverteilung: gegeben

Gruppenunterschiede (zweiseitig): **Nein**





Ergebnisse: App-Kompetenz (TAEG; [11])

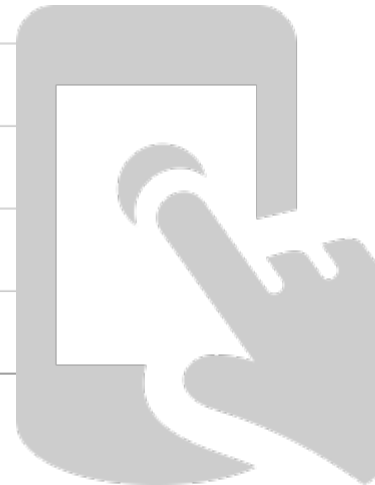
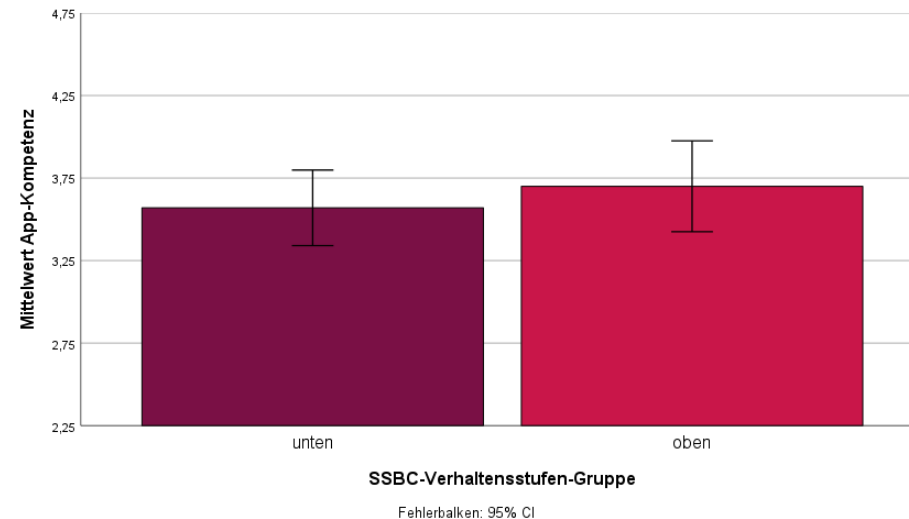
Mittlere Zustimmung: $MW = 3.64$ (= „**trifft eher zu**“; $SD = .09$; Min = 2.25; Max = 4.50)

Vergleich mit den Mittelwerten der Stichprobe ($N = 460$, $MW = 3.47$): signifikant verschieden

Unsere Versuchspersonen schätzen hinsichtlich ihrer **Kompetenz im Umgang mit Apps höher** als die Normstichprobe.

Normalverteilung: gegeben

Gruppenunterschiede (zweiseitig): **Nein**

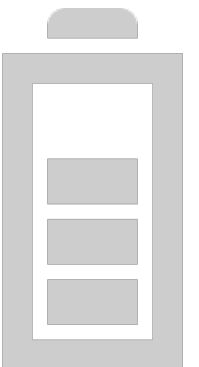




Ergebnisse: Aufgabe 1 Leistung

Informieren über das Risiko einer App

- Beschreiben Sie die Risikofaktoren dieser App in eigenen Worten (5min)
 - Zuordnung von Punkten durch zwei Rater bei Erwähnung von Risiken durch Drittanbieter, Standorterkennung, Zugriff auf die SD-Karte (Zugriff auf persönliche Daten), ID-Übermittlung (Werbung), Sicherheitslücken, Weiterleitung der Daten in unsichere Länder
- Zugewiesenen Punkte wurden zu einem Score verrechnet
- Bei Abweichungen zwischen den Erstkodierenden entschied eine dritte kodierende Person
- Cohens Kappa: 1, Aufgabe $\kappa = .75$ (= substantielle Übereinstimmung, [12])





Ergebnisse: Aufgabe 1 Leistung

Informieren über das Risiko einer App

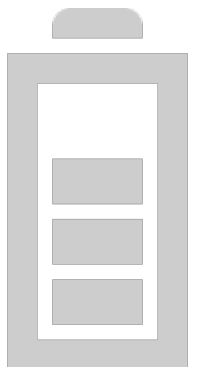
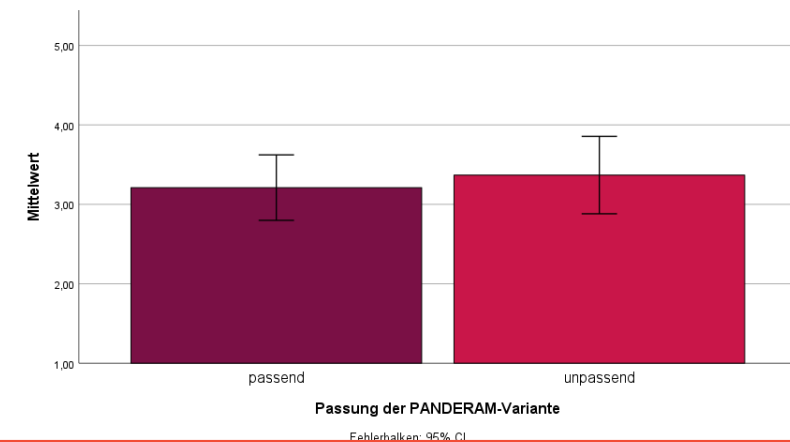
Deskriptive Statistik: $MW = 3.29$ ($SD = .93$, $Min = 1.00$, $Max = 5.00$)

Passende Variante: $n = 19$, $MW = 3.21$ ($SD = .85$, $Min = 2.00$, $Max = 5.00$)

Unpassende Variante: $n = 19$, $MW = 3.37$ ($SD = 1.01$, $Min = 1.00$, $Max = 5.00$)

Keine Annahme der Normalverteilung

Einseitiger Mann Whitney-U-Test: **Kein Unterschied zwischen der passenden und unpassenden Bedingung.**

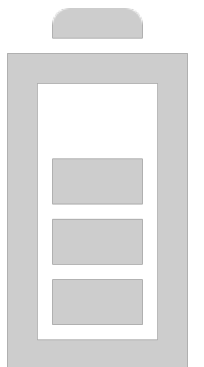




Ergebnisse: Aufgabe 2 Leistung

Beschriebene Maßnahmen zum Absenken des Risikos

- Beschreiben Sie die von Ihnen ergriffenen Maßnahmen mit eigenen Worten (5min).
 - Zuordnung von Punkten, wenn Berechtigungen entzogen wurden, Teilnehmende Apps deinstallieren bzw. Alternativ-Apps installieren würden, die Hintergrundnutzung deaktiviert wurde
- Zugewiesenen Punkte wurden zu einem Score verrechnet
- Bei Abweichungen zwischen den Ratern entschied ein dritter Rater über den endgültigen Punktescore
- Cohens Kappa: 2. Aufgabe $\kappa = .94$ (= fast perfekte Übereinstimmung, [12])





Ergebnisse: Aufgabe 2 Leistung

Beschriebene Maßnahmen zum Absenken des Risikos

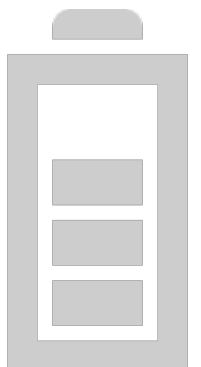
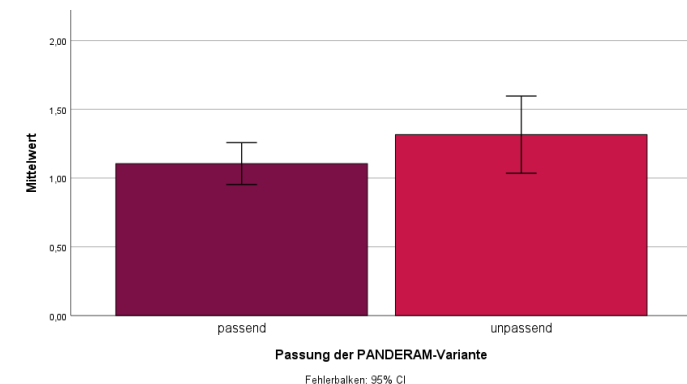
Deskriptive Statistik: $MW = 1.21$ ($SD = .47$, $Min = 1.00$, $Max = 3.00$)

Passende Variante: $n = 19$, $MW = 1.11$ ($SD = .32$, $Min = 1.00$, $Max = 2.00$)

Unpassende Variante: $n = 19$, $MW = 1.32$ ($SD = .58$, $Min = 1.00$, $Max = 3.00$)

Keine Annahme der Normalverteilung

Einseitiger Mann Whitney-U-Test: **Kein Unterschied zwischen der passenden und unpassenden Bedingung.**





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

Mittlerer SUS-Score: $MW = 80.00$ ($SD = 13.90$, $Min = 30$, $Max = 97.50$); entspricht [14] **Note A-**

Testung gegen Benchmark [14]: SUS-Score im unteren „guten Bereich“ (Score 72.6, untere Grenze von B+) ergab signifikanten Unterschied

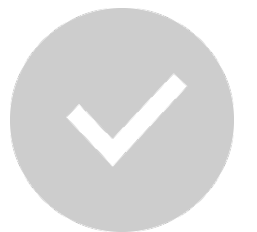
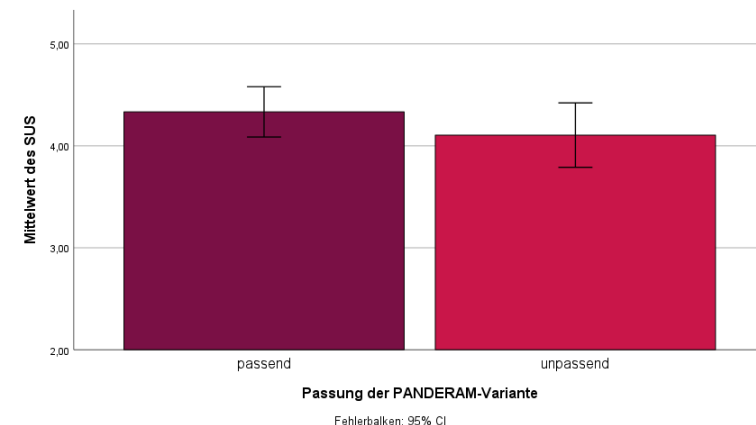
Die Bewertung des PANDERAM-Prototypen hinsichtlich **Usability** fällt insgesamt **sehr gut** aus.

Passende Bedingung: $n = 19$, $MW = 4.33$, $SD = .51$, $Mdn = 4.56$,
 $Min = 2.78$, $Max = 5.00$

Unpassende Bedingung: $n = 19$, $MW = 4.11$, $SD = .66$, $Mdn = 4.11$,
 $Min = 2.11$, $Max = 5.00$

Keine Annahme der Normalverteilung

Einseitiger Mann-Whitney-U-Test: **Kein Unterschied zwischen der passenden und unpassenden Bedingung**





Ergebnisse: Usability 2/4 (PET-USES; [15])

Skala	<i>n</i>	<i>MW</i>	<i>SD</i>	Min	Max	MW AndProtect (<i>N</i> = 22)	Einseitiger t-Test vs. MW 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	$t(37) = 2.12, p = .021, d = .34$

Die Skalen wurden alle **positiv** („stimmt“) beantwortet und weisen signifikant **höhere Werte im Vergleich zur AndProtect App** auf.





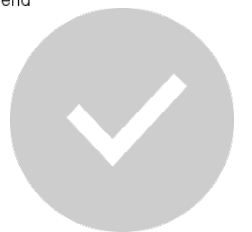
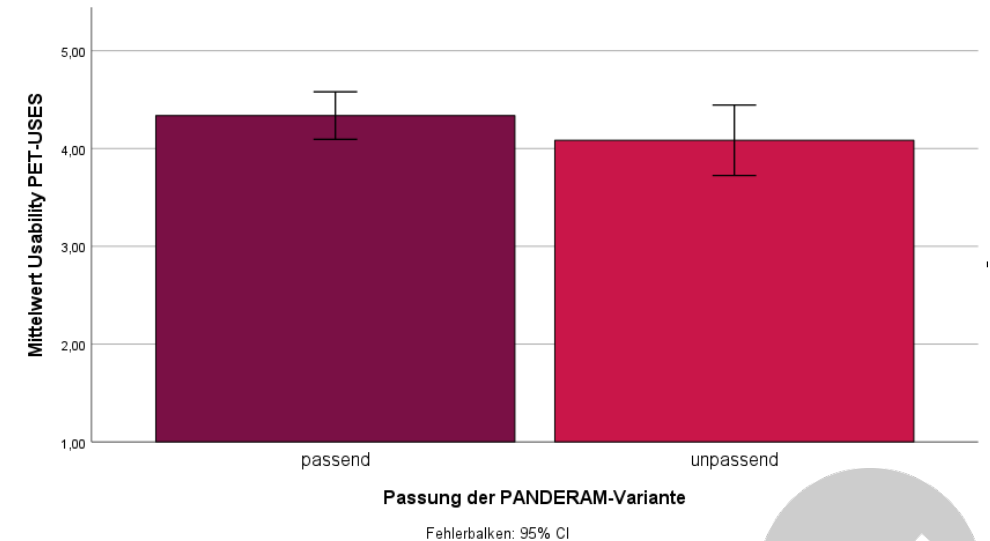
Ergebnisse: Usability 3/4 (PET-USES; [15])

Passende Bedingung: $n = 19$, $MW = 4.34$, $SD = .50$, $Min = 3.10$,
 $Max = 5.00$

Unpassende Bedingung: $n = 19$, $MW = 4.08$, $SD = .75$, $Min = 2.10$,
 $Max = 5.00$

Annahme der Normalverteilung

Einseitiger t-Test für unabhängige Stichproben: **Kein
Unterschied zwischen der passenden und unpassenden
Variante.**





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

Deskriptive Statistik für die Subskala **Usability**: $MW = 5.72$ ($SD = .99$, $Min = 3.00$, $Max = 7.00$).

Einseitiger t-Test gegen Skalenwert 5: **Eher positive Einschätzung der Usability.**

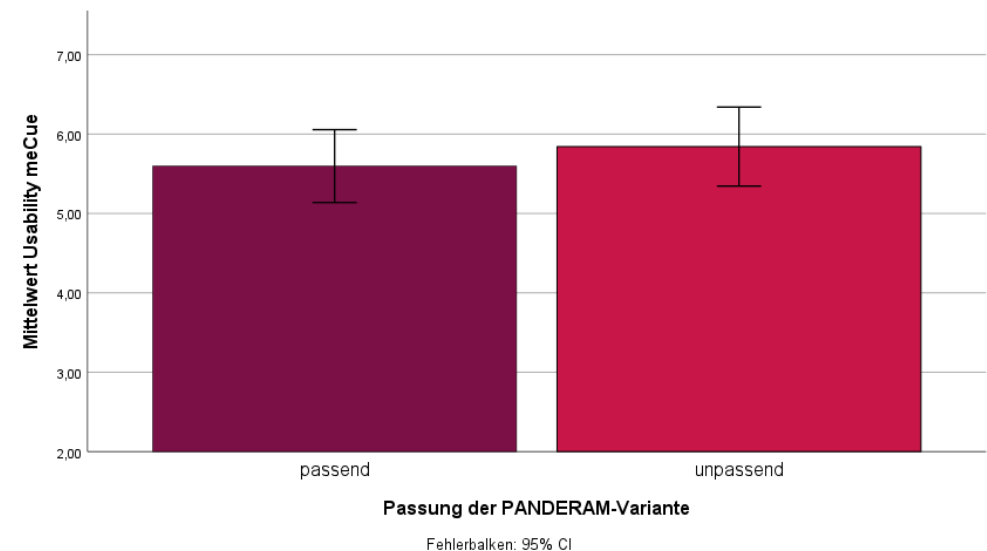


Passende Bedingung: $n = 19$, $MW = 5.60$, $SD = .95$, $Mdn = 5.67$,
 $Min = 3.00$, $Max = 7.00$

Unpassende Bedingung: $n = 19$, $MW = 5.84$, $SD = 1.03$, $Mdn = 6.00$,
 $Min = 3.33$, $Max = 7.00$

Keine Annahme der Normalverteilung

Einseitiger Mann-Whitney-U-Test: **Kein Unterschied zwischen der passenden und unpassenden Variante**





Ergebnisse: User Experience 1/4 (UEQ; [17])

Skala	<i>n</i>	<i>MW</i>	<i>SD</i>	Min	Max	Grenzwert Benchmark	Einseitige Testung gegen unterer Wert Bewertungskategorie	Bewertung
Attraktivität	38	1.49	.72	-.83	3	1.17-1.52	$t(37) = 2.74, p = .005, d = .44$	Über dem Durchschnitt
Durchschaubarkeit	38	1.88	1.01	-2.25	3	1.56-1.9	$t(37) = 1.92, p = .032, d = 3.1$	Gut
Vorhersagbarkeit	38	1.80	1.80	.50	2.75	=> 1.65	$t(37) = 1.63, p = .056$	Gut
Stimulation	38	1.13	.77	2.50	3.25	.99-1.31	$t(37) = 1.08, p = .143$	Unter Durchschnitt
Originalität	38	1.07	.89	1.50	2.75	1.05-1.4	$t(37) = .155, p = .439$	Über Durchschnitt
Effizienz	38	1.72	.65	-.25	3	1.47-1.78	$t(37) = 2.41, p = .011, d = .39$	Gut

UX wird überwiegend als „gut“ bzw. „über dem Durchschnitt“ bewertet, nur **Stimulation** wird **unterdurchschnittlich** eingeschätzt.





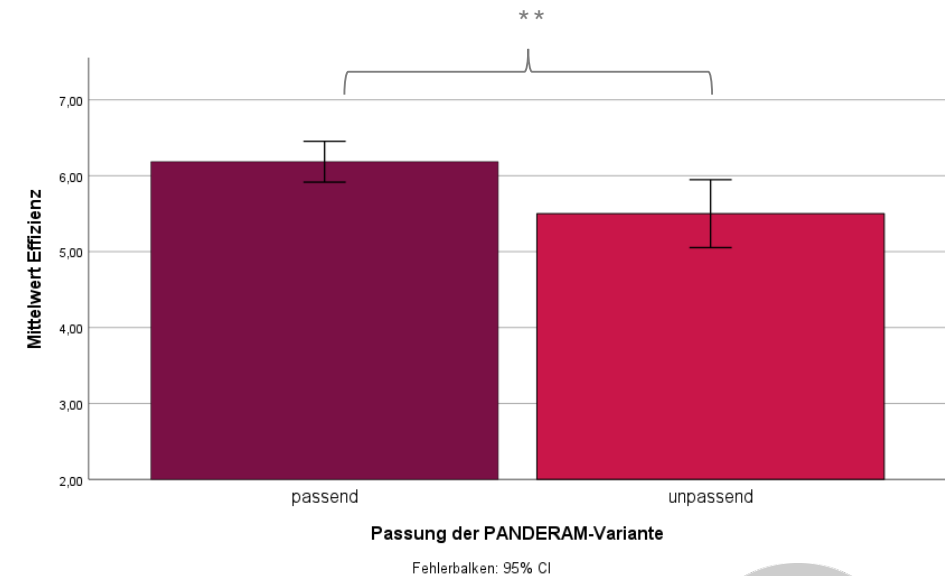
Ergebnisse: User Experience 2/4_(UEQ; [17])

Passende Bedingung: $n = 19$, $MW = 6.18$, $SD = .56$, $Mdn = 6.00$,
Min = 5.00, Max = 7.00

Unpassende Bedingung: $n = 19$, $MW = 5.50$, $SD = .93$, $Mdn = 5.50$,
Min = 3.50, Max = 7.00

Keine Annahme der Normalverteilung

Einseitiger Mann-Whitney-U-Test: **Es wurde in der Subskala Effizienz ein Unterschied zwischen der passenden und unpassenden Bedingung gefunden.**





Ergebnisse: User Experience 3/4 (meCUE 2.0; [16])

Deskriptive Statistik für die Subskala **positive Emotionen**: $MW = 3.94$ ($SD = .91$, Min = 1.5, Max = 5.33).

Einseitiger Test gegen Skalenwert 4: nicht signifikant.

Neutrale Einschätzung auf der Subskala „positive Emotionen“ (weder/noch).

Deskriptive Statistik für die Subskala **negative Emotionen**: $MW = 2.37$ ($SD = .1.03$, Min = 1, Max = 5.83).

Einseitiger Test gegen Skalenwert 3: signifikanter Unterschied.

Ablehnung für die Subskala „negative Emotionen“.

**Teilnehmende erleben keine negativen Emotionen bei der Nutzung der App.
Das Erleben positiver Emotionen wird als neutral beschrieben.**





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

Positive Emotionen

Passende Bedingung: $n = 19$, $MW = 4.10$, $SD = .75$, $Min = 2.50$, $Max = 5.33$

Unpassende Bedingung: $n = 19$, $MW = 3.79$, $SD = 1.05$, $Min = 1.50$, $Max = 5.33$

Annahme der Normalverteilung

Einseitiger t-Test für unabhängige Stichproben: **Kein Unterschied zwischen der passenden und unpassenden Variante gefunden.**

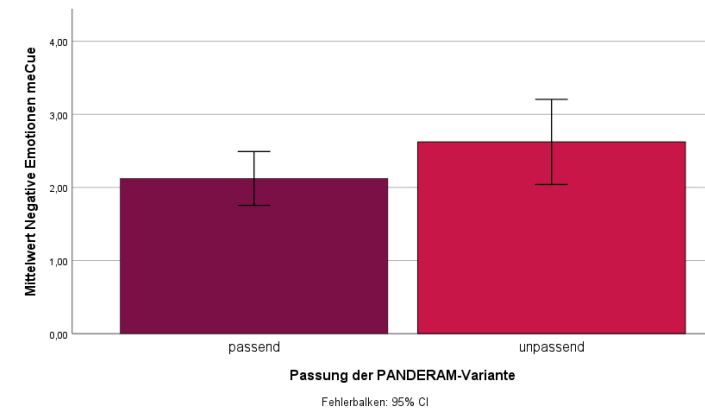
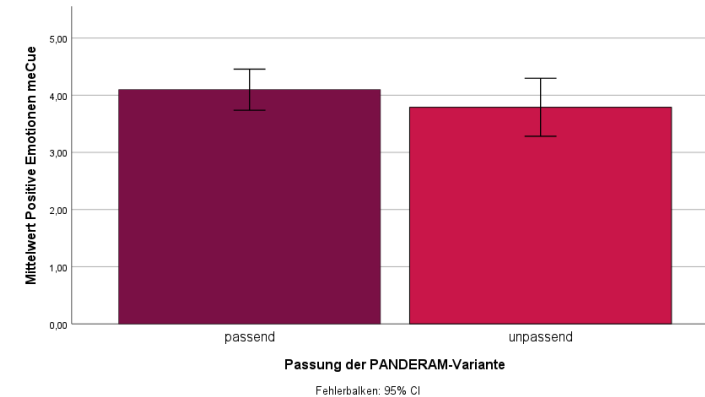
Negative Emotionen

Passende Bedingung: $n = 19$, $MW = 2.12$, $SD = .76$, $Min = 1.00$, $Max = 3.50$

Unpassende Bedingung: $n = 19$, $MW = 2.62$, $SD = 1.21$, $Min = 1.00$, $Max = 5.83$

Annahme der Normalverteilung

Einseitiger t-Test für unabhängige Stichproben: **Kein Unterschied zwischen der passenden und unpassenden Variante gefunden.**





Ergebnisse: Transparenz (SIPAS; [18])

Im Vergleich zum Skalenmittelwert (3,5) ergeben sich für Vorhersagbarkeit ein signifikanter Unterschied (kleiner Effekt) und bei Transparenz und Verständlichkeit signifikante Unterschiede (große Effekte).

SIPAS - Subskala	T-Wert	Df	Signifikanz (einseitig)	Effektstärke Cohens d
Transparenz	9.50	37	<.001	1.54
Verständlichkeit	6.12	37	<.001	.99
Vorhersagbarkeit	1.92	37	.031	.31

Die **PANDERAM-App** wird als überdurchschnittlich **transparent, verständlich und vorhersagbar** bewertet.



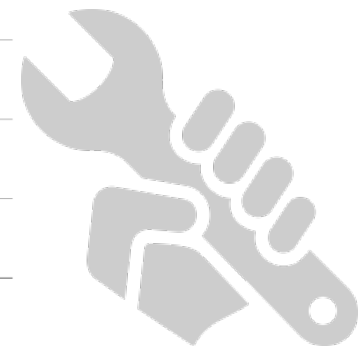
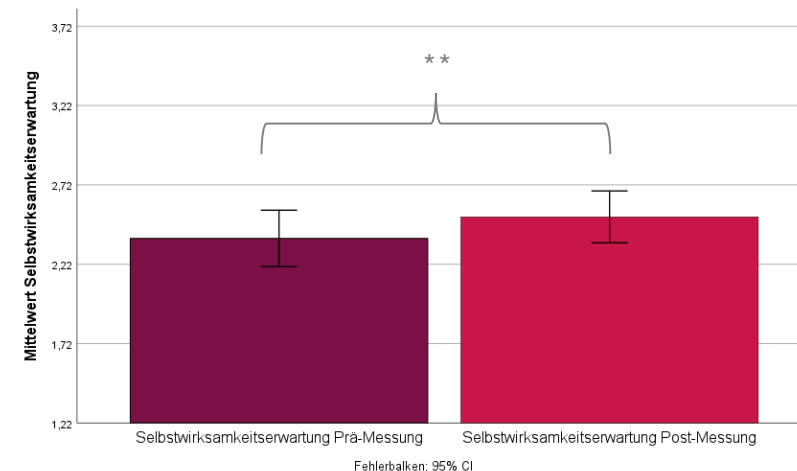


Ergebnisse: Selbstwirksamkeitserwartung 1/2 [19]

Messung	<i>n</i>	<i>MW</i>	<i>SD</i>	MIN	MAX
Prä-Messung	38	2.38	.54	1.22	3.44
Post-Messung	38	2.52	.50	1.30	3.70

T-Test (einseitig) Prä- und Postmessung:
signifikanter Unterschied.

Teilnehmende berichten nach Vorstellung und Interaktion mit der PANDERAM-App **höhere Selbstwirksamkeitserwartung**, d.h. sie fühlen sich nachher eher befähigt bei Datenschutz- und Privatsphäre Problemen agieren zu können.



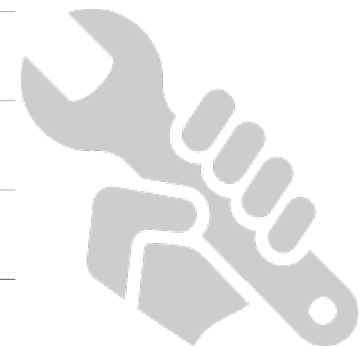
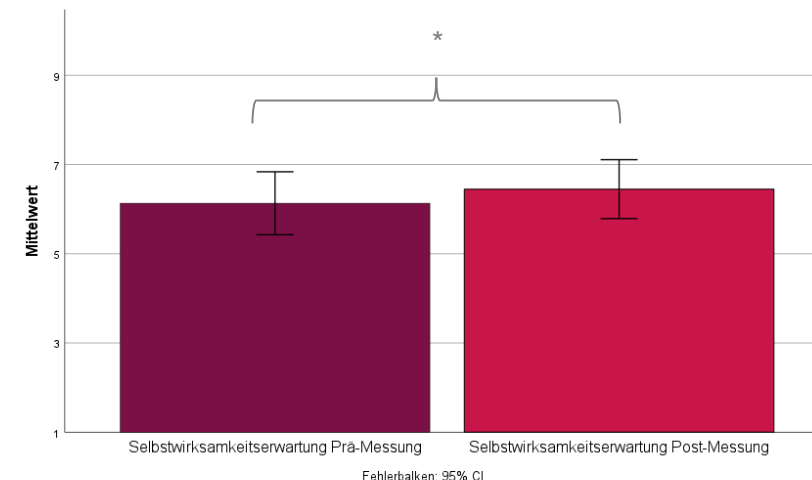


Ergebnisse: Selbstwirksamkeitserwartung 2/2 [20]

Messung	<i>n</i>	<i>MW</i>	<i>SD</i>	MIN	MAX
Prä-Messung	38	6.13	2.15	1	10
Post-Messung	38	6.45	2.01	3	10

T-Test (einseitig) Prä- und Postmessung:
signifikanter Unterschied.

Teilnehmende berichten auch hier nach Präsentation und Interaktion mit der PANDERAM-App eine **höhere Selbstwirksamkeitserwartung** bzgl. Datenschutz- und Privatsphäre Problemen zu haben.





Ergebnisse: Nutzungsintention (UTAUT; [21])

Messung	<i>n</i>	<i>MW</i>	<i>SD</i>	MIN	MAX	Einseitige Testung gegen den Skalenmittelwert
Leistungserwartung	38	3.20	1.12	1.00	5.00	$t(37) = -4.38, p < .001, d = -.71$
Anstrengungserwartung:	38	5.58	1.25	1.00	7.00	$t(37) = 7.81, p < .001, d = 1.27$
Einstellung gegenüber Technologie	38	4.72	.81	1.75	6.25	$t(37) = 5.49, p < .001, d = .89$
Sozialer Einfluss	38	4.91	.93	3.00	6.75	$t(37) = 6.07, p < .001, d = .99$
Unterstützende Bedingungen	38	5.18	.98	1.67	6.67	$t(37) = 7.38, p < .001, d = 1.20$
Selbstwirksamkeitserwartung	38	4.95	1.24	1.25	7.00	$t(37) = 4.75, p < .001, d = .77$
Ängstlichkeit	38	2.39	1.14	1.00	5.75	$t(37) = -8.70, p < .001, d = -1.41$
Nutzungsintention	38	5.09	1.11	1.00	7.00	$t(37) = 6.03, p < .001, d = .98$

Die **Leistungserwartung** der App fällt **negativ** aus. Die **anderen Aspekte** der App (z.B. Erwartung, wie anstrengend die App ist oder Ängstlichkeit) werden **positiv bewertet**.





Ergebnisse: Lob für die App (qualitativ)

Welche Aspekte der PANDERAM-App haben Ihnen **besonders gut gefallen**?
Nennen Sie die drei wichtigsten.

Es wurden **121 Aussagen** in Kategorien mit zwei parallelen Ebenen kodiert: 1. Ebene = Unterscheidung zwischen „Informationen“ (37% der Aussagen), „Gestaltung“ (35%), und „Handlungsoptionen“ (28%).

2. Ebene: Detailliertere Inhalte (siehe nächste Folie)

Intercoder-Reliabilität (gewichtet):

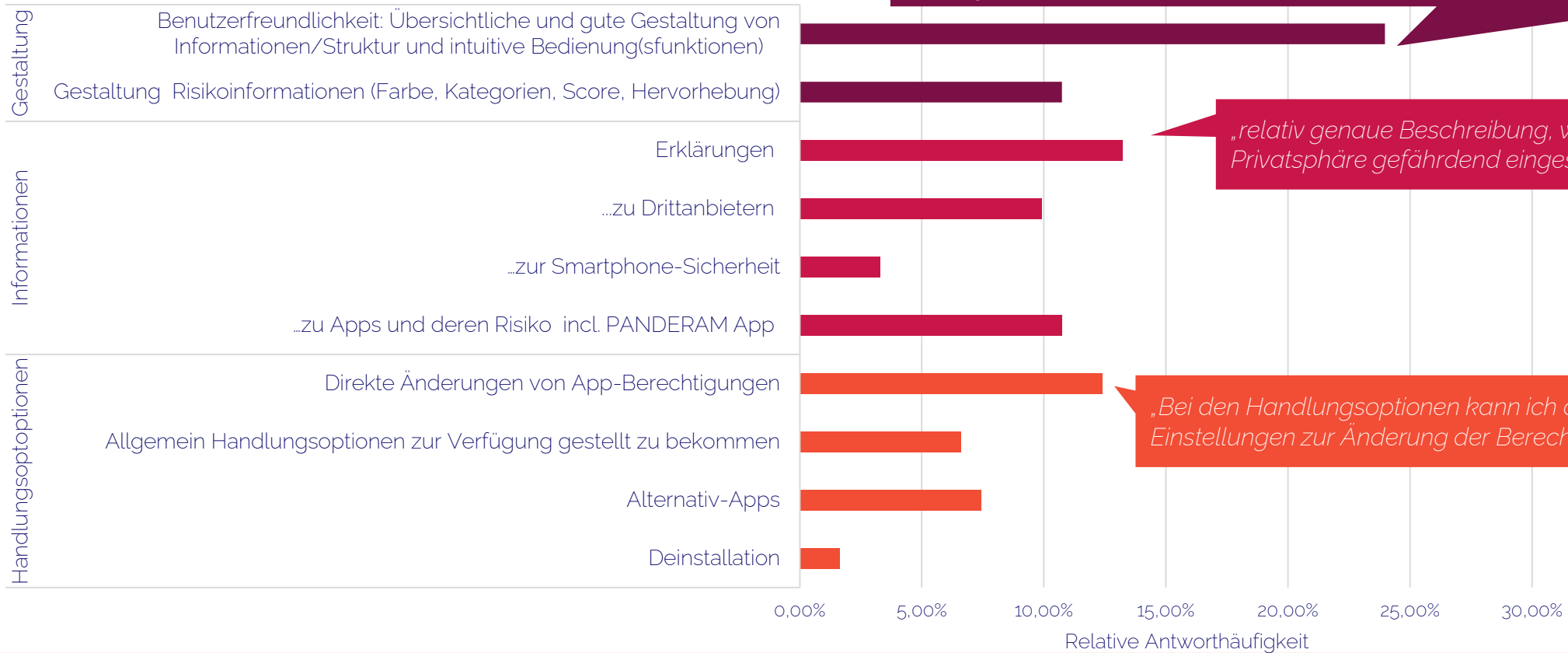
$\kappa = 0.84$ (= (fast) perfekte Übereinstimmung in der Kodierung; [12])





Ergebnisse: Lob für die App (qualitativ)

„Die Benutzeroberfläche war sehr verständlich und angenehm gestaltet. Man hat sofort gewusst, wo man Informationen finden kann, welche man gerade sucht.“



„relativ genaue Beschreibung, warum etwas als Privatsphäre gefährdend eingestuft wird.“

„Bei den Handlungsoptionen kann ich direkt zu den Einstellungen zur Änderung der Berechtigungen kommen [...]“





Ergebnisse: Kritik für die App (qualitativ)

Welche **Verbesserungsmöglichkeiten** sehen Sie für die PANDERAM-App?
Nennen Sie die drei wichtigsten Dinge.

Es wurden **93 Aussagen** in Kategorien mit zwei parallelen Ebenen kodiert: 1. Ebene = Unterscheidung zwischen „**Informations- und Funktionsumfang der App ausbauen**“ (44% der Aussagen), „**Gestaltung vorhandener Informationen verbessern**“ (43%). 9% der Antworten entfielen auf „keine“ Verbesserungsvorschläge.

2. Ebene: Detailliertere Inhalte (siehe nächste Folie)

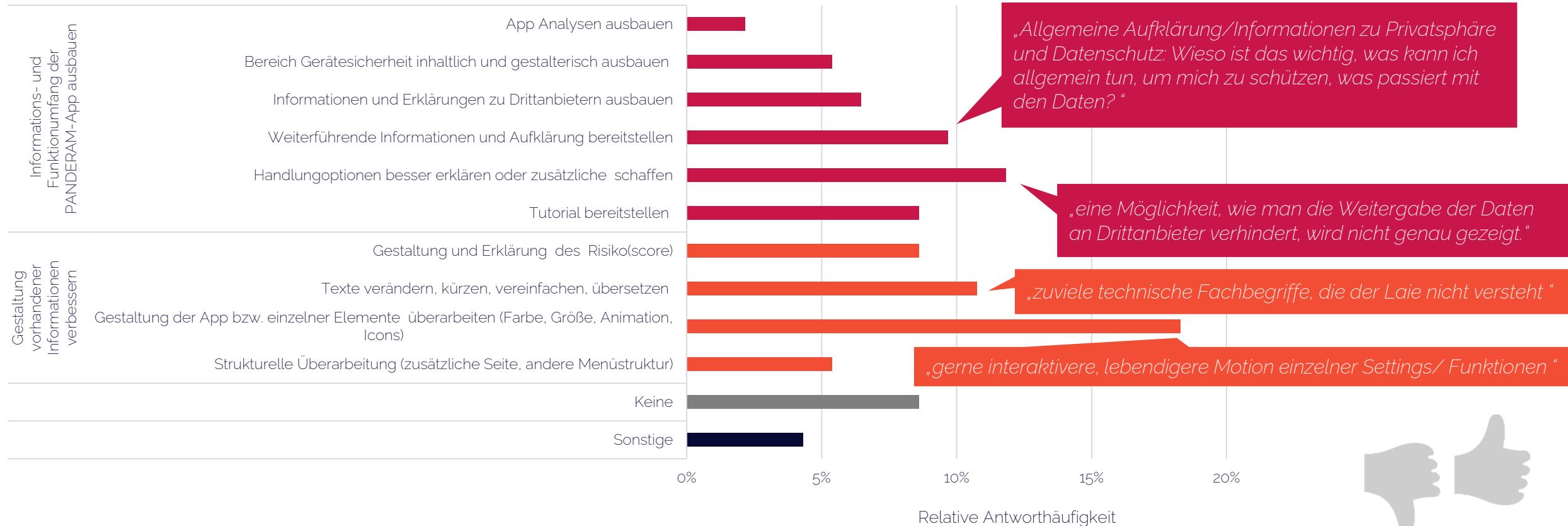
Intercoder-Reliabilität (gewichtet):

$\kappa = 0.65$ (= substantielle Übereinstimmung; [12])





Ergebnisse: Kritik für die App (qualitativ)





Ergebnisse: Schulnote

Welche Schulnote würden Sie der PANDERAM-App insgesamt geben?

Auf einer Skala von (1 = *sehr gut*, 2 = *gut*, 3 = *befriedigend*, 4 = *ausreichend*, 5 = *mangelhaft* und 6 = *ungenügend*) vergaben die $N = 38$ Teilnehmenden des Laborversuches im Mittel die Note „**gut**“ ($Mdn = 2.00$; $Min = 1.00$; $Max = 5.00$).

Die beiden **Verhaltensstufen-Gruppen unterscheiden sich nicht** in ihrer Bewertung.
Die Versuchsbedingung macht für die Benotung ebenfalls keinen Unterschied.





Ergebnisse: Vergleich App-Isomorphie Verfahren



Ausgangs-App: AirBrush

Ranking/ Verfahren		VAE		UMAP	
Platz 1		PicShot	↔	Ultimate Photo Mixer	
		↕		↕	
Platz 2		Photo Editor Pro	↔	Foto Blender	





Ergebnisse: Vergleich App-Isomorphie Verfahren

Welchen der beiden Vorschläge empfinden Sie als ähnlicher zur Original-App?

Die überwiegende Mehrheit (86.8%; 33) der Probanden urteilte das **PicShot (VAE1) ähnlicher** zu AirBrush ist als Ultimate Photo Mixer (UMAP 1).



Je nach vorheriger Auswahl sollten die Probanden nachfolgend beantworten:

1.) Sofern [PicShot/Ultimate Foto Mixer] weniger riskant für meine Privatsphäre ist, würde ich die Original-App AirBrush mit dieser Alternativ-App ersetzen. (1 = *stimme gar nicht zu*, 2 = *stimme nicht zu*, 3 = *stimme zu*, 4 = *stimme völlig zu*)

- Antworten für PicShot (VAE 1, $n = 33$): $MW = 2,58$ ($SD = 0,90$, Min = 1, Max = 4)
- Antworten für Ultimate Foto Mixer (UMAP 1, $n = 5$): $MW = 2,00$ ($SD = 0,00$, Min = 2, Max = 2)

Da die Gruppen sehr ungleich besetzt sind, können hier keine Vergleiche gerechnet werden.





Ergebnisse: Vergleich App-Isomorphie Verfahren

Zudem wurden die Probanden gebeten für beide Apps separat eine Ähnlichkeits-einschätzungen zwischen (1 = *gar nicht ähnlich* bis 10 = *sehr ähnlich abzugeben*).

Ähnlichkeit PicShot (VAE 1) zu AirBrush: $MW = 5.53$ ($SD = 2.04$, Min = 1, Max = 9)

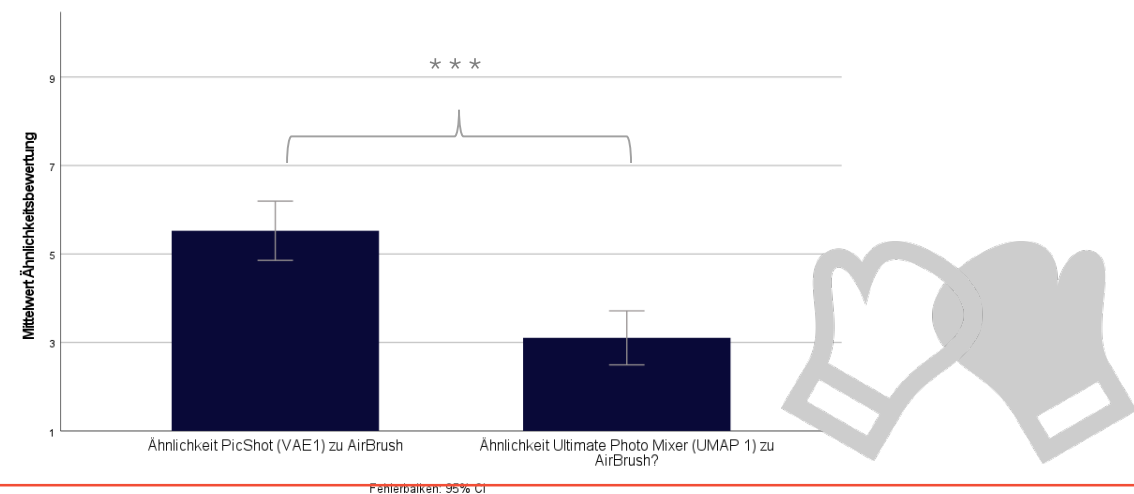
Ähnlichkeit Ultimate Photo Mixer (UMAP 1) zu AirBrush: $MW = 3.11$ ($SD = 1.86$, Min = 1, Max = 7)



Diese Einschätzungen sind jeweils nicht normalverteilt.

Die Ähnlichkeitsbewertung von **PicShot und Ultimate Photo Mixer unterscheiden sich signifikant**.

Dabei wird **PicShot** ($Mdn = 6.00$) **als deutlich ähnlicher zu AirBrush** ($Mdn = 2.50$) **bewertet**, als Ultimate Foto Mixer.





Ergebnisse: Vergleich App-Isomorphie Verfahren

Welchen der beiden Vorschläge empfinden Sie als ähnlicher zur Original-App?

Die überwiegende Mehrheit (84.2%; 32) der Probanden urteilte, dass **Photo Editor Pro (VAE2) ähnlicher** zu AirBrush ist als Foto Blender.



Je nach vorheriger Auswahl sollten die Probanden nachfolgend wieder beantworten:

Sofern [Photo Editor Pro/Foto Blender] weniger riskant für meine Privatsphäre ist, würde ich die Original-App AirBrush mit dieser Alternativ-App ersetzen. (1 = *stimme gar nicht zu*, 2 = *stimme nicht zu*, 3 = *stimme zu*, 4 = *stimme völlig zu*)

- Antworten für Photo Editor Pro (VAE 2, $n = 32$): $MW = 2.81$ ($SD = 0.90$, Min = 1, Max = 4)
- Antworten für Foto Blender (UMAP 2, $n = 6$): $MW = 3.17$ ($SD = 1.33$, Min = 1, Max = 4)

Da die Gruppen sehr ungleich besetzt sind, können hier keine Vergleiche gerechnet werden.





Ergebnisse: Vergleich App-Isomorphie Verfahren

Zudem wurden die Probanden gebeten für beide Apps separat eine Ähnlichkeits-einschätzung zugeben (1 = *gar nicht ähnlich* bis 10 = *sehr ähnlich*).

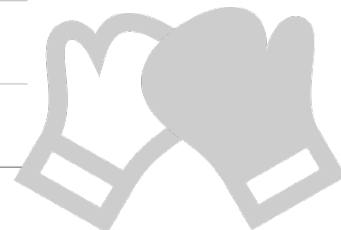
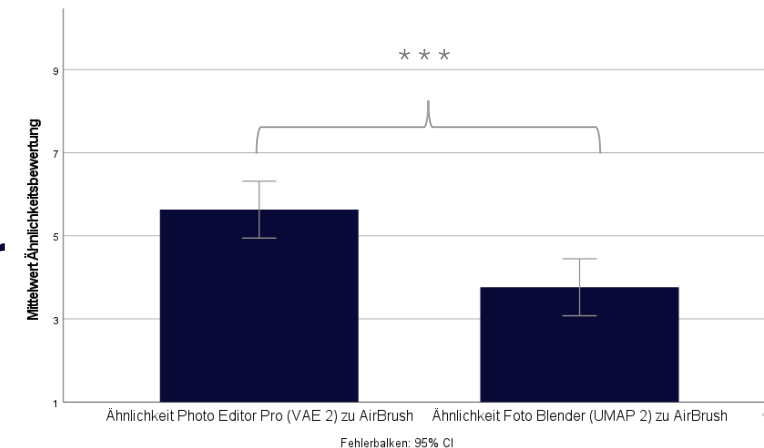
Ähnlichkeit Photo Editor Pro (VAE 2) zu AirBrush: $MW = 5.63$ ($SD = 2.09$, Min = 2, Max = 9)

Ähnlichkeit Foto Blender (UMAP 2) zu AirBrush: $MW = 3.76$ ($SD = 2.09$, Min = 1, Max = 9)



Die Einschätzungen sind jeweils nicht normalverteilt.

Die Ähnlichkeitsbewertungen von **Photo Editor Pro** und **Foto Blender** unterscheiden sich statistisch signifikant. Dabei wird **Photo Editor Pro** ($Mdn = 6.00$) **deutlich ähnlicher** als Foto Blender ($Mdn = 3.00$) bewertet.





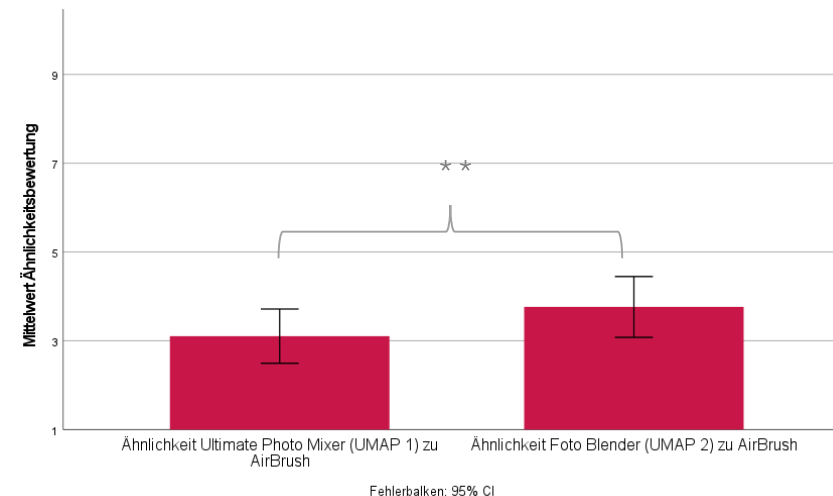
Ergebnisse: Vergleich App-Isomorphie Verfahren

Vergleich Ähnlichkeitsbewertungen VAE

Die Ähnlichkeitsbewertungen von **PicShot** (VAE 1, $Mdn = 6.00$) und **Photo Editor Pro** (VAE 2, $Mdn = 6.00$) **unterscheiden sich nicht** statistisch signifikant.

Vergleich Ähnlichkeitsbewertungen UMAP

Die Ähnlichkeitsbewertungen von **Ultimate Photo Mixer** (UMAP 1, $Mdn = 2.50$) und **Foto Blender** (UMAP 2, $Mdn = 3.00$) **unterscheiden sich** statistisch signifikant. Der **zweite Vorschlag** wird hierbei als **geringfügig ähnlicher bewertet als der erste Vorschlag**, welchen das Verfahren liefert.





Zusammenfassung

App-Bewertung

- Die **Usability** der PANDERAM-App wird als positiv bzw. „**sehr gut**“ eingestuft
- Die **User Experience** bewerten die Teilnehmenden als „über dem Durchschnitt“ bzw. „**gut**“
 - Allerdings wird die Stimulation der App von Nutzenden, als Schwäche bewertet
 - Das Erleben positiver Emotionen wird als neutral eingestuft, das Erleben von negativen Emotionen hingegen wird abgelehnt
- Die **Transparenz** der PANDERAM-App wird als **überdurchschnittlich** angegeben
- Die **PANDERAM-App erhöht die Selbstwirksamkeitserwartung** der Teilnehmenden, mit Datenschutzproblemen umgehen zu können
- Die **Nutzungsintention** fällt insgesamt **positiv** aus, allerdings gehen Teilnehmende nicht davon aus, dass die App ihre (Arbeits-)Leistung verbessern kann

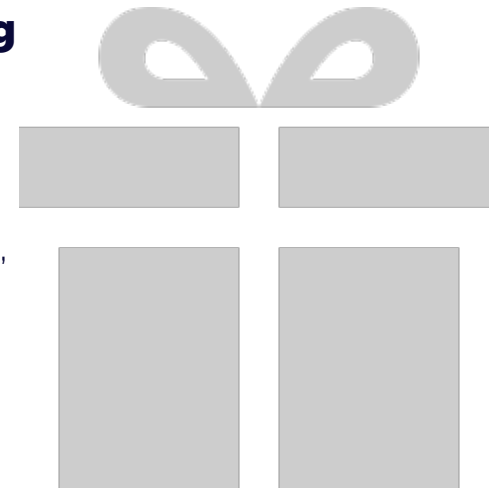




Zusammenfassung

App-Bewertung

- Am häufigsten wünschten sich die Teilnehmenden:
 - **Besser erklärte und mehr Handlungsmöglichkeiten**, um bspw. die **Weitergabe der Daten an Drittanbietende verhindern** zu können
 - mehr allgemeine **Informationen und Aufklärung zum Thema** und ein **Tutorial** für die Erstnutzung
 - Zudem eine Überarbeitung einzelner Elemente z.B. eine **interaktivere Gestaltung** einzelner Settings/Funktionen wird gewünscht
 - Es wurden die verwendeten Texte z.B. hinsichtlich **Fachbegriffen** kritisiert.
- Häufigstes Lob: App - die Benutzeroberfläche wurde als **benutzerfreundlich** bewertet, **Erklärungen** und **direkte Handlungsoption** wurden ebenfalls gelobt
- Die App wird allgemein mit der **Schulnote „gut“** bewertet





Zusammenfassung

Vergleich passende vs. unpassende App Variante

- Ob die Teilnehmenden einen für ihre Verhaltensstufe passenden bzw. unpassenden Prototyp präsentiert bekamen, spielt für die Usability und Leistung keine Rolle
- Bei den Subskalen der User Experience unterscheidet sich nur die Subskala „**Effizienz**“ zwischen den Teilnehmenden, die eine passende bzw. unpassende Variante erhalten
 - Das heißt, dass die App für die Teilnehmenden in der unpassenden Bedingung als ineffizienter, unpragmatischer bzw. überladener wahrgenommen wurde, als für die Teilnehmenden in der passenden Bedingung





Zusammenfassung

Laborversuch App-Isomorphie

- Bei beiden Urteilen, die die Teilnehmenden fällen sollten, entschieden sie sich für die **vorgeschlagene App des VAE Verfahrens** und beurteilten diese als **ähnlicher zur Original-App**
- Bei den beiden Ähnlichkeitsbewertungen schätzten die Teilnehmenden den **Vorschlag des VAE Verfahrens signifikant ähnlicher** ein **als** den Vorschlag des **UMAP** Verfahrens
- Vergleicht man den jeweils 1. und 2. Vorschlag eines Verfahrens, ergaben sich beim VAE Verfahren keine Unterschiede, beim UMAP Verfahren wurde der 2. Vorschlag, den das Verfahren liefert als geringfügig ähnlicher bewertet als der 1. Vorschlag.

Das VAE-Verfahren liefert aus Sicht der Teilnehmenden bessere Vorschläge.





Ausblick

- Anhand der Ergebnisse soll der PANDERAM-Prototyp überarbeitet werden
- Die neue Version wird in einem kleineren Laborversuch erneut getestet
- Hier liegt der Fokus auf dem Sicherheitsbereich der App





Vielen Dank für die Aufmerksamkeit!

Susen Döbelt

Wilhelm-Raabe-Str. 43
09120 Chemnitz

Telefon: 0371 531 33615

E-Mail: susen.doebelt@psychologie.tu-chemnitz.de