

HCI Meilenstein 4

Elektronisches Curriculum

Team 1

Pascal Attwenger, Philipp Hiermann, Sandra Markhart

15. Juni 2014

1 Einleitung

Im Verlauf der letzten Wochen wurde in Gruppenarbeit ein neues Interface für die Prüfungsleistungen und das Mitteilungsblatt („eCurriculum“) entwickelt. Ziel dieser Studie ist es zu überprüfen, ob unser neues Interface wirklich besser zu bedienen ist als das alte Interface.

Im Folgenden wird zunächst der Studienaufbau beschrieben. Wir gehen auf die zu belegenden Hypothesen ein sowie auf deren Quantisierung und die gewählte Testmethodik.

Anschließend werden die durch die Studie gefundenen Daten zunächst mittels deskriptiver Statistik dargelegt und beschrieben; hier werden auch erste mögliche Folgerungen aufgezeigt. Diese gefundenen Daten werden daraufhin mit Methoden der Inferenzstatistik auf ihre Signifikanz untersucht und aufgrund dieser Erkenntnisse diskutiert.

Abschließend werden noch Reflexionen über das Gesamtprojekt hinsichtlich seiner Kernerkenntnisse sowie persönliche Anmerkungen jedes Teammitglieds festgehalten.

2 Studienaufbau

Die Studie vergleicht unser neues Interface mit dem alten Interface (univis/Mitteilungsblatt). Die Aufgaben werden zuerst mit dem jeweils alten System durchgeführt und in Folge dessen in Hinblick auf die Bedienbarkeit mit dem neuen System verglichen.

Zusätzlich werden Fragebögen nach Vorbild von AttrackDiff verwendet, um die subjektive Zufriedenheit der Benutzer mit den Systemen festzuhalten.

2.1 Hypothesen

Nullhypothese 1: User können mit unserem neu erstellten Interface Aufgaben schlechter oder gleich gut lösen als mit dem alten Interface.

Alternativhypothese 1: User können mit unserem neu erstellten Interface Aufgaben besser lösen als mit dem alten Interface.

Nullhypothese 2: User sind mit unserem neu erstellten Interface weniger zufrieden oder gleich zufrieden als mit dem alten Interface.

Alternativhypothese 2: User sind mit unserem neu erstellten Interface zufriedener als mit dem alten Interface.

2.2 Testbare Aufgaben

Aufgabe 1: Welche Lehrveranstaltungen können im Modul “Modul VMI Vertiefung Medieninformatik” absolviert werden? (Mitteilungsblatt)

Aufgabe 2: Wieviele ECTS bringt die UE Arbeitstechniken Multimediajournalismus? (Mitteilungsblatt)

Aufgabe 3: Welche Note habe ich in der Übung “UE Technische Grundlagen und Systemsoftware”? (univis)

2.3 Variablen

2.3.1 Performanzmetriken

- Zeit in Sekunden
- Anzahl der Klicks bei den univis-Aufgaben (deprecated)

2.3.2 Subjektives Nutzerfeedback

Ein AttrakDiff-Fragebogen, bei dem folgende Eigenschaften mit 7 verschiedenen Möglichkeiten zu bewerten sind:

- einfach – kompliziert
- menschlich – technisch
- verständlich – unverständlich
- übersichtlich – unübersichtlich
- innovativ – konservativ
- kreativ – phantasielos
- schön – hässlich
- fröhlich – deprimierend

Datei auf cewebs: Meilenstein 4 – Team 1 – Fragebogen

2.4 Pre-Test

In Folge des Pre-Tests mussten wir feststellen, dass die Variable “Klicks” nicht zielführend war, da in den meisten Fällen nur ein Klick benötigt wurde, weshalb diese in der späteren tatsächlichen Studie weggelassen wurde.

Die restlichen Daten waren jedoch hinsichtlich ihrer Eignung für die Studie nicht beeinflusst und konnten daher weiterverwendet werden.

2.5 Testpersonen

RL: männlich, 21, studiert Rechtswissenschaften an der JKU Linz

JM: männlich, 21, studiert Rechtswissenschaften an der JKU Linz

MS: männlich, 21, studiert Informatik an der JKU Linz

JF: weiblich, 14, Schülerin

NM: weiblich, 47, noch nie mit dem Univis gearbeitet

DK: männlich, 27, studiert Medieninformatik an der Uni Wien

JB: weiblich, 30, studierte 4 Semester Psychologie an der Uni Wien

PH: männlich, 27, studiert Medieninformatik an der Uni Wien

2.6 Methodik

Bei der Auswahl der Testpersonen wurde darauf geachtet, eine ausgewogene Mischung verschiedenster Studienhintergründe zu erreichen: Es wurden sowohl Studierende der Uni Wien als auch die anderer Universitäten interviewt; zwei der Probandinnen hatten überhaupt keine Vorerfahrungen mit Universitäts-IS-Systemen.

Den Testpersonen wurden zuerst die zwei Aufgaben zum Mitteilungsblatt gestellt und dabei die Zeit gestoppt, daraufhin wurden sie gebeten, ihre Eindrücke vom Mitteilungsblatt im AttrackDiff-Fragebogen festzuhalten. Mit den Aufgaben zum Univis und zum neu entwickelten eCurriculum wurde in Folge gleich verfahren.

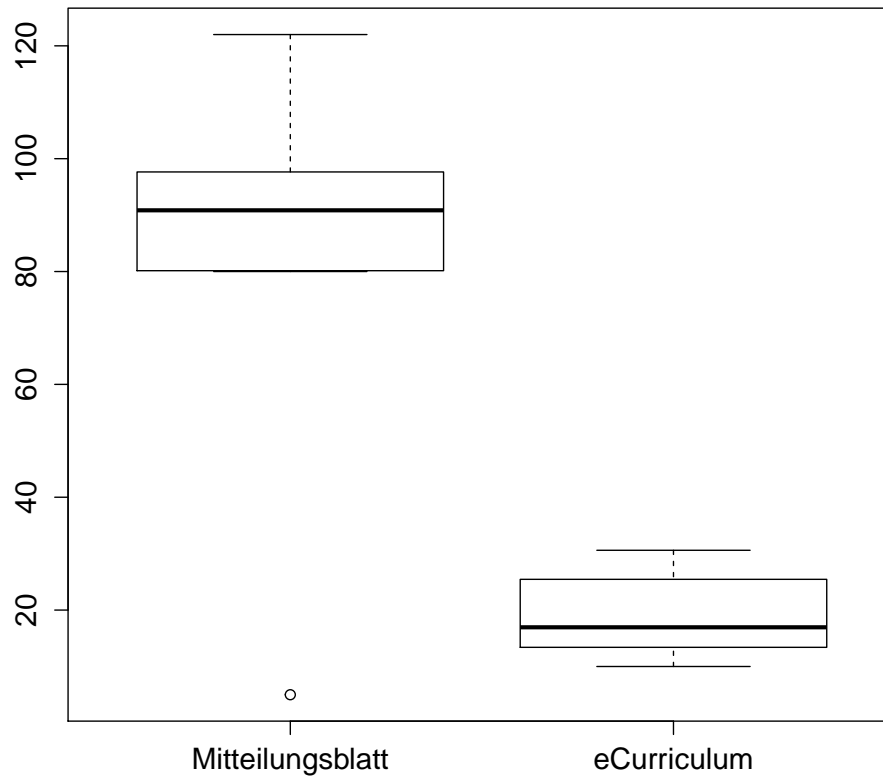
3 Resultate

3.1 Deskriptive Statistik

Zeit in Sekunden

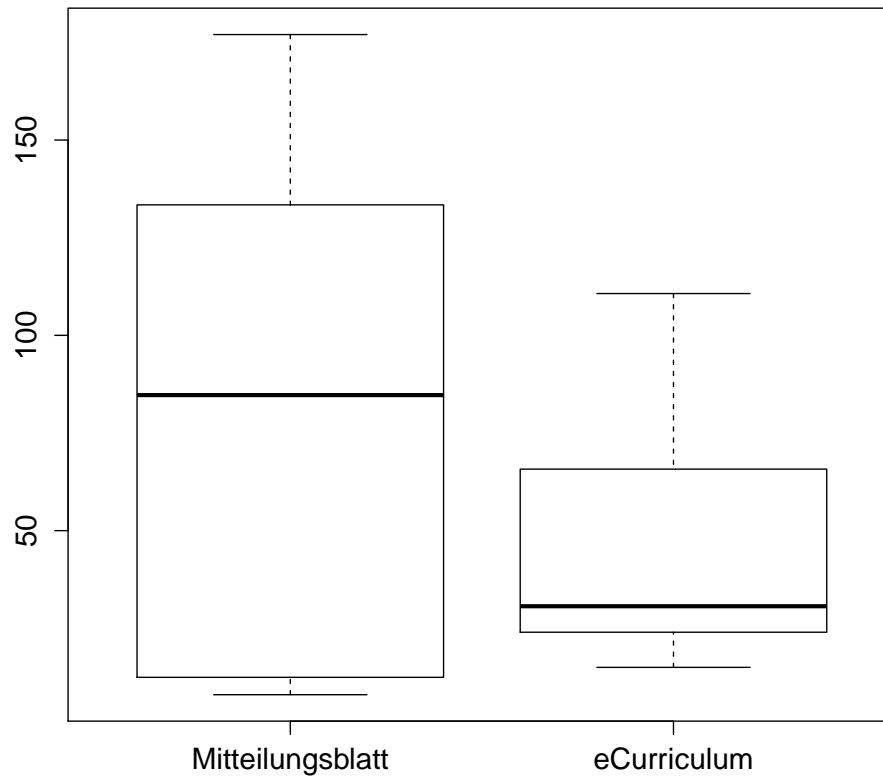
Name	Mitteilungsblatt/Univis			eCurriculum		
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Aufgabe 1	Aufgabe 2	Aufgabe 3
RL	80,3	133,8	65,3	30,6	29,7	27,6
JM	80	80,4	42,3	25,9	80,5	14,4
MS	98,3	13,4	28,5	16,9	110,7	13,5
JF	91,7	11,5	84	16,8	31,6	33,3
NM	90	89	148	10	51	19
DK	5	8	32	10	15	16
JB	122	177	47	17	29	32
PH	97	133	20	25	19	48

Zeit in Sekunden für die 1. Aufgabe



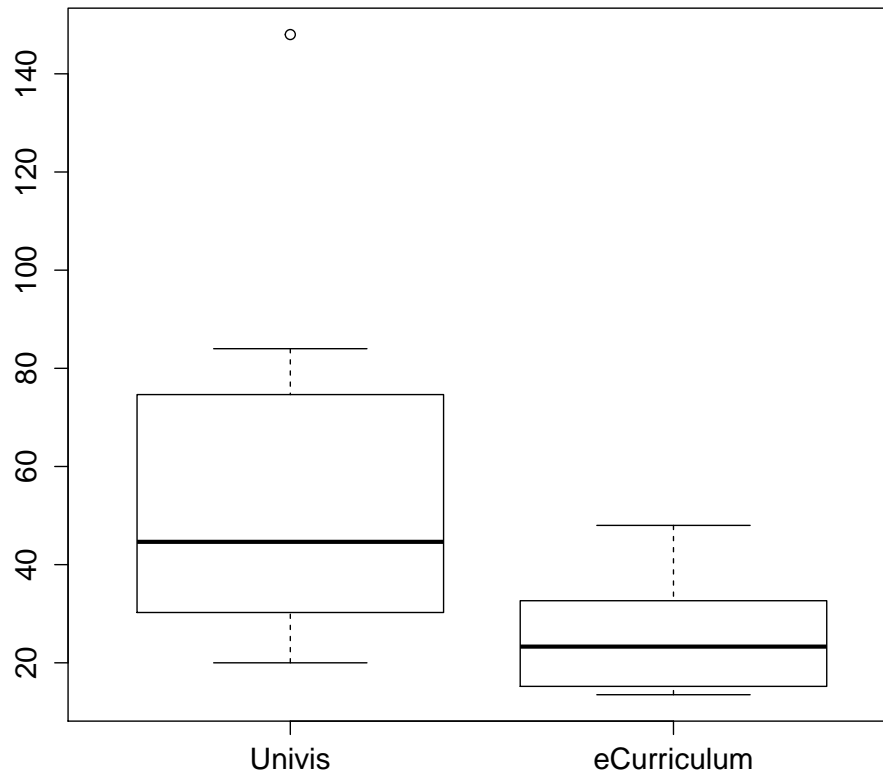
Bei der ersten Aufgabe (Finden der LVs zum Modul VMI) zeigt sich bei Betrachtung des Boxplots über die benötigte Zeit ein deutliches Bild, demzufolge die Bearbeitung der Aufgabe mit dem alten Mitteilungsblatt deutlich mehr Zeit in Anspruch nimmt als mit dem neuen eCurriculum. Ein einziger Ausreißer der Mitteilungsblatt-Zeiten liegt unter allen des eCurriculums, alle anderen Versuchspersonen waren mit dem neuen System schneller.

Zeit in Sekunden für die 2. Aufgabe



Bei der zweiten Aufgabe (Finden der ECTS-Punkte einer Lehrveranstaltung) zeigt sich bei den Versuchen mit dem Mitteilungsblatt eine deutlich größere Streuung als beim eCurriculum. Zwar liegt der Median beim neuen System deutlich unter dem alten, jedoch gab es auch einige Personen, die mit dem alten System schneller waren.

Zeit in Sekunden für die 3. Aufgabe



Bei der dritten Aufgabe wurde das Finden einer einzelnen Prüfungsleistung zwischen Univis und eCurriculum verglichen. Hier zeigt sich beim neuen System, dass bei durchschnittlich weniger Bearbeitungszeit auch die Streuung geringer war, viele Versuchspersonen also ähnlich wenig Zeit benötigten.

Beim Univis liegen zwar auch Werte vor, die ähnlich schnell wie im eCurriculum waren, allerdings ist die Streuung bedeutend größer, und es kam auch zu einem sehr auffälligen Ausreißer.

Subjektives Nutzerfeedback

Für die Erstellung der Barplots wurden zu jedem AttrackDiff-Begriffspaar die durchschnittlichen Bewertungen über alle Datensätze (auf einer Skala von 1 bis 7) erstellt. Zu beachten ist hierbei, dass niedrigere Werte die bessere Bewertung bedeuten.

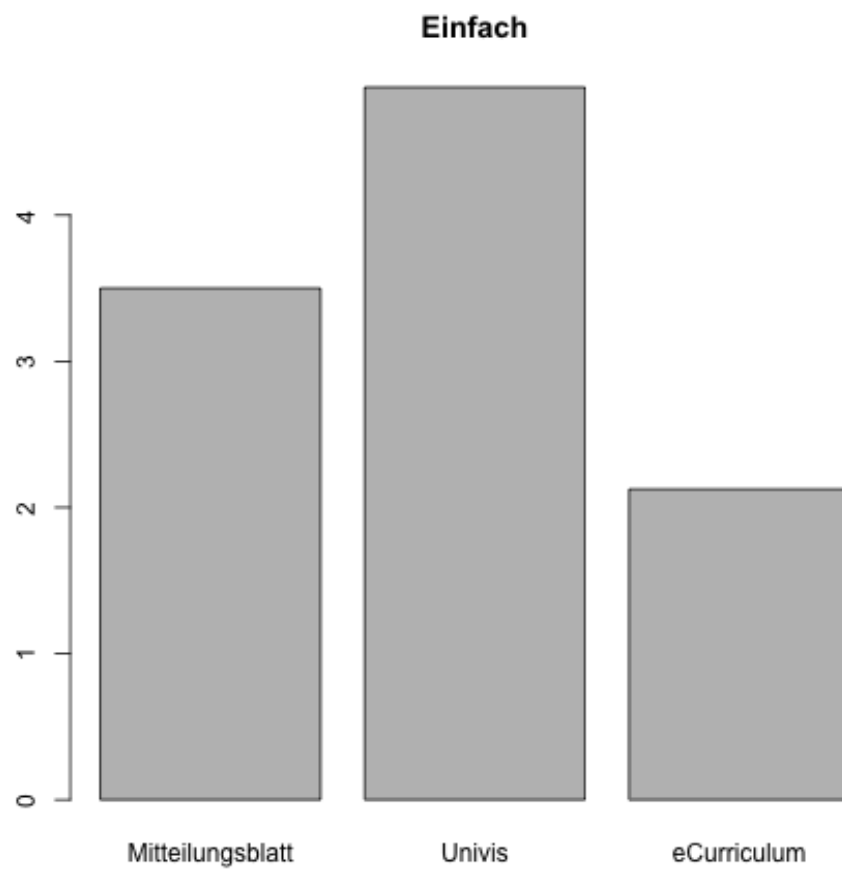
Bei jedem erfassten Aspekt zeigt sich, dass das von uns neu entwickelte System die besseren Bewertungen erhielt, weshalb die Vermutung besteht, dass die User damit tendenziell zufriedener sind.

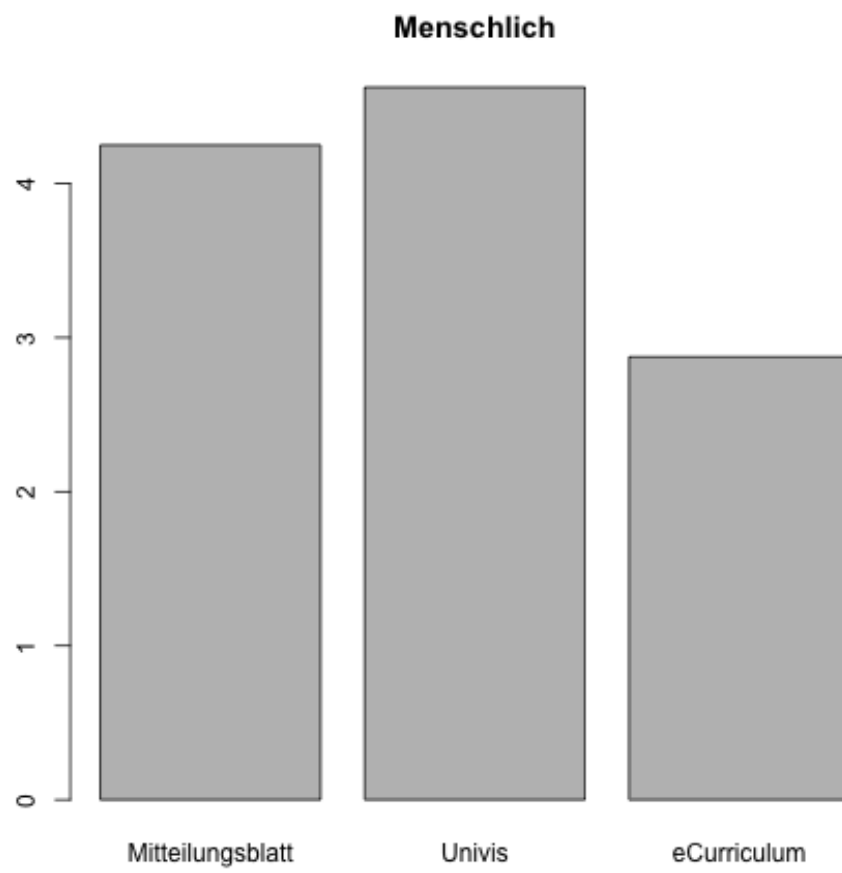
Die Untersuchung dieser Ergebnisse hinsichtlich Signifikanz findet sich im nächsten Abschnitt.

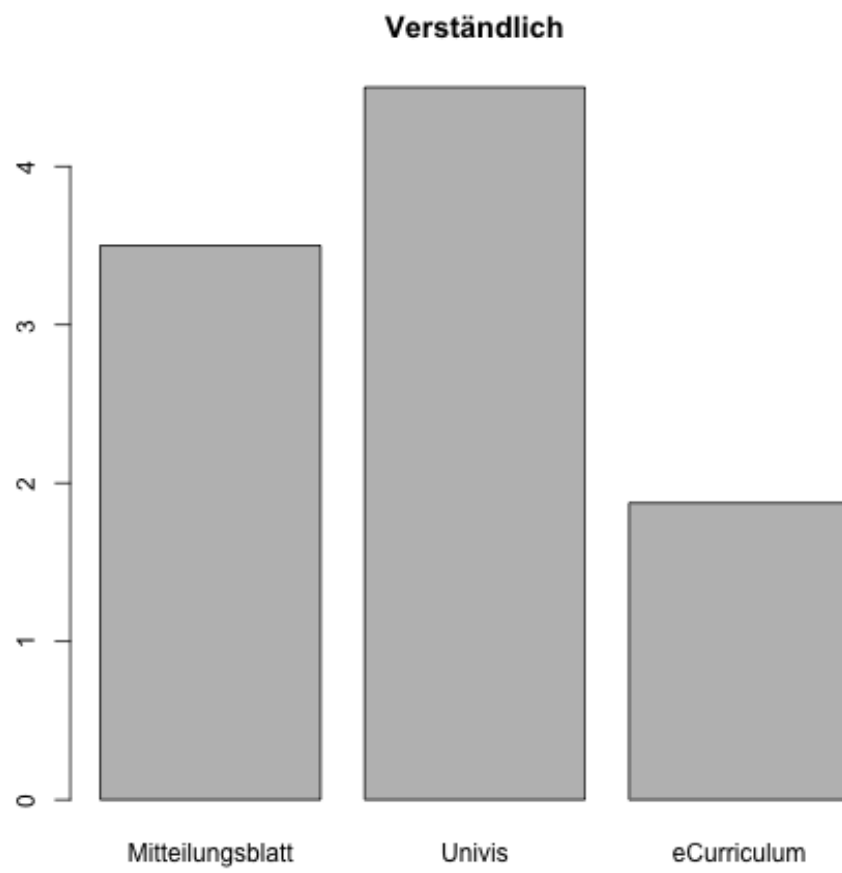
	Mitteilungsblatt							
Name	einfach	menschlich	verständlich	übersichtlich	innovativ	kreativ	schön	fröhlich
RL	2	2	1	7	7	6	7	6
JM	5	4	5	6	4	5	6	6
MS	5	4	3	6	6	7	5	4
JF	3	6	3	4	5	5	4	6
NM	5	4	3	6	6	6	6	5
DK	2	6	3	6	7	7	5	4
JB	4	4	5	3	5	6	7	4
PH	2	4	5	1	4	5	7	4

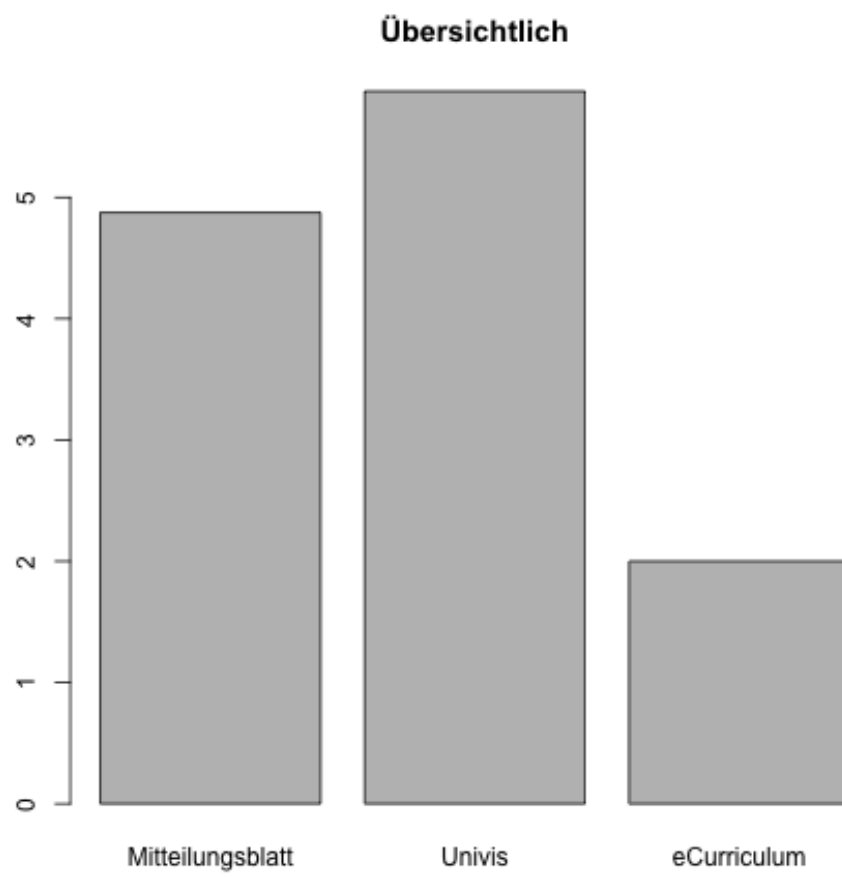
	Univis							
Name	einfach	menschlich	verständlich	übersichtlich	innovativ	kreativ	schön	fröhlich
RL	2	2	1	7	7	6	7	6
JM	5	4	5	6	4	5	6	6
MS	5	4	3	6	6	7	5	4
JF	3	6	3	4	5	5	4	6
NM	5	4	3	6	6	6	6	5
DK	2	6	3	6	7	7	5	4
JB	4	4	5	3	5	6	7	4
PH	2	4	5	1	4	5	7	4

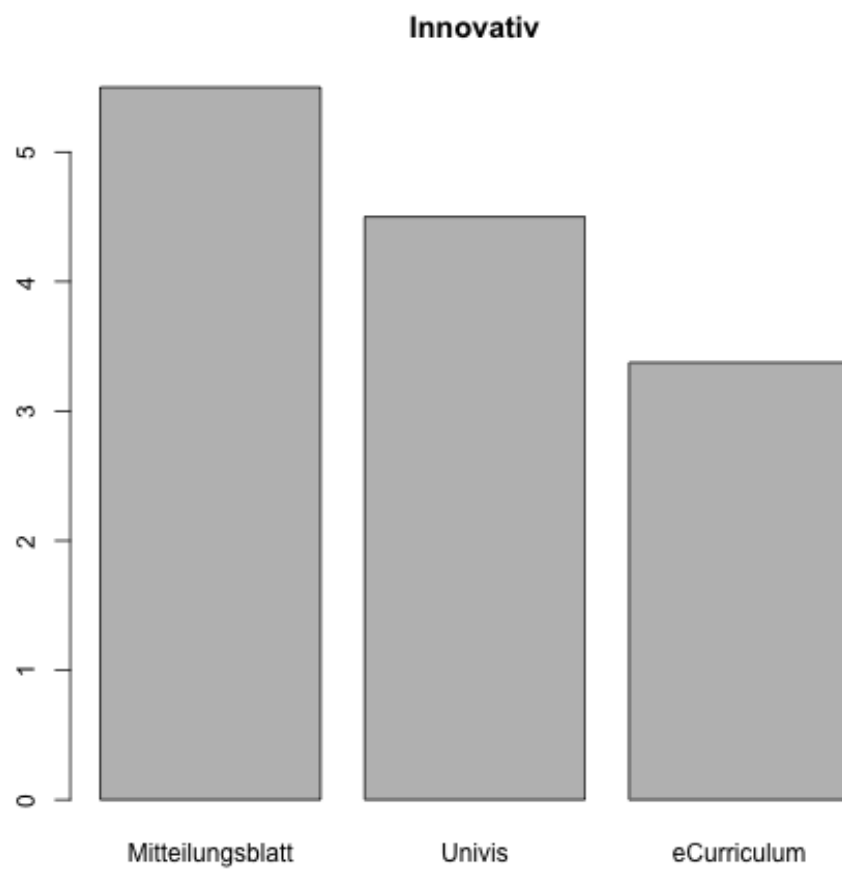
	eCurriculum							
Name	einfach	menschlich	verständlich	übersichtlich	innovativ	kreativ	schön	fröhlich
RL	1	2	1	2	3	4	3	3
JM	3	4	3	3	4	4	4	4
MS	3	3	2	1	2	2	2	2
JF	1	2	2	1	1	2	2	2
NM	3	3	2	3	4	4	2	4
DK	2	3	2	2	3	2	3	3
JB	1	3	2	2	5	3	3	4
PH	3	3	1	2	5	2	4	1

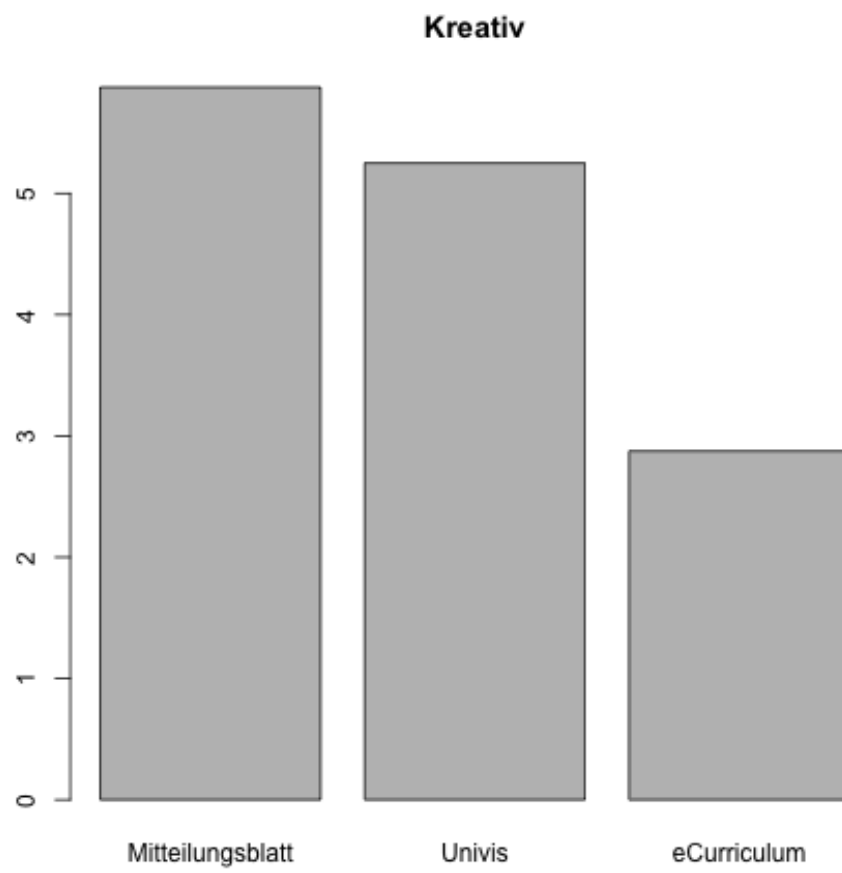


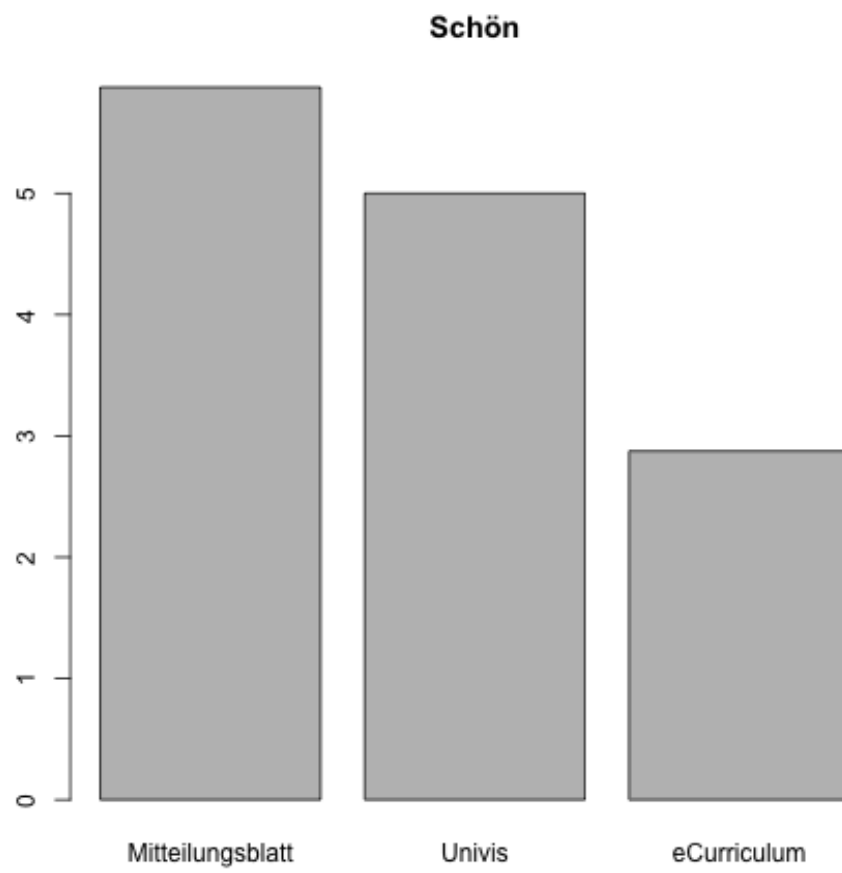


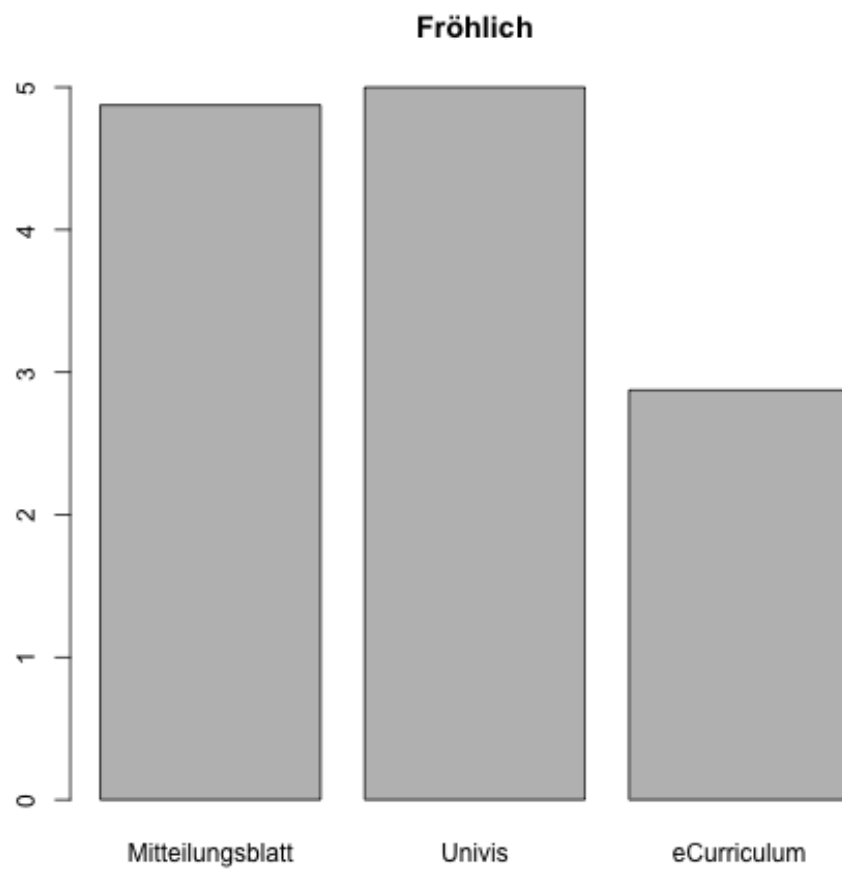












3.2 Inferenzstatistik

Die inferenzstatistischen Untersuchungen wurden in R mittels t-Test durchgeführt. Verglichen wurden zu den 3 Aufgaben jeweils die Samples des alten und neuen Systems.

Für die Untersuchung der Zufriedenheit wurden die Mittelwerte über alle subjektiven Bewertungen der alten Systeme (Mitteilungsblatt und Univis) einerseits und des neuen eCurriculums andererseits gebildet, und auf Basis dieser ein t-Test durchgeführt.

Insbesondere wurden die p-Values als Wert der Fehlerwahrscheinlichkeit, dass die Nullhypothese zu Unrecht verworfen wurde, betrachtet.

Zeit in Sekunden

Aufgabe 1:

Welch Two Sample t-test

```
data: a1_mb and a1_ec
t = 5.1747, df = 7.672, p-value = 0.0004839
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 40.88108      Inf
sample estimates:
mean of x mean of y
 83.0375  19.0250
```

Aufgabe 2:

Welch Two Sample t-test

```
data: a2_mb and a2_ec
t = 1.3532, df = 10.48, p-value = 0.1022
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-11.64372      Inf
sample estimates:
mean of x mean of y
 80.7625  45.8125
```

Aufgabe 3:

Welch Two Sample t-test

```
data: a3_uv and a3_ec
t = 2.1438, df = 8.156, p-value = 0.03187
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
```

```
4.434431    Inf
sample estimates:
mean of x mean of y
58.3875    25.4750
```

Zufriedenheit

Welch Two Sample t-test

```
data: meansAlt and meansNeu
t = 6.1818, df = 12.889, p-value = 1.721e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.61058    Inf
sample estimates:
mean of x mean of y
4.867188 2.609375
```

4 Diskussion

4.1 Alternativhypothese 1: User können mit unserem neu erstellten Interface Aufgaben besser lösen als mit dem alten Interface.

Man kann anhand der Zeiten in Sekunden in der tabellarischen Ansicht und anhand der Boxplots schon erkennen, dass die Aufgaben mit dem neuen Interface von den Usern schneller gelöst wurden, als mit dem alten Interface. Jedoch kann man hier noch nichts über die Signifikanz aussagen, weshalb wir hier jeweils zu jeder Aufgabe noch einen t-Test zu den Zeiten in Sekunden durchgeführt haben.

Nachdem wir einen Two-Sampled-t-Test durchgeführt haben und uns die p-Value angesehen haben, können wir bei der 1. Aufgabe sagen, dass bei einem Signifikanzniveau von 95% die 1. Aufgabe mit dem neuem Interface besser bzw. schneller gelöst werden kann als mit dem alten Interface, da die Nullhypothese mit einer Fehlerwahrscheinlichkeit von 0,0004839 verworfen werden kann.

Bei der 2. Aufgabe gibt es mit einer p-Value von 0,1022 zwar einen deutlich erkennbaren aber bei 95-prozentigem Signifikanzniveau nicht signifikanten Unterschied. Die Nullhypothese – das neue System ist für die Bearbeitung dieser Aufgabe schlechter oder nur gleich gut als das alte – muss hier bei gegebener Datenlage also vorerst beibehalten werden.

Bei der 3. Aufgabe kann die Nullhypothese mit einer p-Value von 0,03187 und bei einem Signifikanzniveau von 95% ebenfalls verworfen werden, weshalb die Alternativhypothese angenommen werden kann und somit die User mit unserem neu erstellten Interface Aufgaben signifikant besser lösen können als mit dem alten Interface.

4.2 Alternativhypothese 2: User sind mit unserem neu erstellten Interface zufriedener als mit dem alten Interface.

Anhand der verschiedenen Barplots ist hier auch wieder ersichtlich, dass das eCurriculum insgesamt als „einfacher“, „menschlicher“, „verständlicher“, „schöner“, usw. wahrgenommen wurde als das Mitteilungsblatt und das Univis; hier ist auch gut ersichtlich, dass das Univis insgesamt als am wenigsten benutzerfreundlich wahrgenommen wurde.

Auch hier wurde erneut ein Two-Sampled-t-Test durchgeführt um etwas über die Signifikanz aussagen zu können. Wurde zuerst der Mittelwert je Eigenschaft berechnet (von „einfach“ bis „fröhlich“) und mit diesen Werten dann der t-Test durchgeführt.

Das Ergebnis erhielten wir eine p-Value von $1,721e-05$ weshalb man hier auch wieder sagen kann, dass die Nullhypothese bei einem Signifikanzniveau von 95% mit einer Fehlerwahrscheinlichkeit von $1,721e-05$ verworfen werden kann. Somit kann auch hier auch wieder unsere Alternativhypothese angenommen werden, welche besagt, dass User mit dem neu erstellten Interface zufriedener sind als mit dem alten Interface.

4.3 Schlussfolgerungen

Insgesamt konnten wir feststellen, dass Aufgaben mit dem neuem System (eCurriculum) stets schneller zu lösen waren, und zwar bis auf eine einzige Ausnahme auf statistisch signifikantem Niveau. Zusätzlich wurde es von den Benutzern als benutzerfreundlicher angesehen wurde als das alte System (Univis, Mitteilungsblatt). Aus diesem Grund können wir behaupten, dass unser System ein Erfolg und besser als das alte System ist.

5 Appendix

Enthält:

- Beantwortete Fragebögen als .pdf oder .jpg
- Daten als .csv-Datei
- Plots als .pdf oder .png
- R-Skripte für die Plots und die T-Tests

Datei auf cewebs: Meilenstein 4 – Team 1 – Appendix

6 Reflexion über das Gesamtprojekt

6.1 Kernerkenntnisse und Designempfehlungen aus dem Projekt

Das erklärte Ziel war es, das eCurriculum übersichtlich zu gestalten. Die Auswertung der Fragebögen lässt darauf schließen, dass uns dies sehr gut gelungen ist. Die tabellarische Form des Studienplans ist gut angekommen und wurde als übersichtlich empfunden. Im aktuellen UNIVIS ist es ja bekanntlich schwierig, sich zurecht zu finden. Sehr gut angekommen ist auch die zusätzliche Funktion, das Curriculum nach Semestern zu gliedern, anstatt nach Modulgruppen. Die Möglichkeit, sich direkt bei der Ansicht einer Lehrveranstaltung für diese anzumelden ist auch sehr empfehlenswert.

6.2 Arbeitsverteilung, Kommunikation, Lernprozess und Zufriedenheit

6.2.1 Sandra Markhart

Meistens trafen wir uns alle 2 Wochen bzw. dann wenn ein neuer Meilenstein anstand für etwa 3,5 Stunden um den Meilenstein zu besprechen und zu bearbeiten. Die Treffen haben wir uns dabei entweder direkt in der Einheit oder per Email ausgemacht. Den 1. Meilenstein haben wir dabei bis auf die Erwartungen jedes einzelnen Teammitglieds zusammen gelöst. Für den 2. Meilenstein trafen wir uns nur für die Erstellung des Low-Fidelity-Prototypen. Der restliche Meilenstein 2 wurde dann in Einzelarbeit, mit Absprache per Mail gelöst. Bei Meilenstein 3 und 4 haben wir die Fragebögen und die Vorgehensweise für die Interviews jeweils in unseren Treffen festgelegt. Die Befragung hat dann jeder für sich geführt. Die Auswertung der Interview bzw. des Experiments wurde dann wieder in einem Treffen durchgeführt.

Zur Arbeitsverteilung ist dabei insgesamt zu sagen, dass viele Aufgaben zusammen gelöst wurden. Der High-Fidelity-Prototyp wurde, bis auf ein paar visuelle Änderungen und Verbesserungsvorschläge von mir erstellt. Bei der Durchführung der Interviews und des Experiments haben zwar alle Teammitglieder beigetragen, jedoch wurden die meisten Befragungen von Pascal durchgeführt. Beschreibung der Prototypen und die Beschreibung und Diskussion der Auswertungen wurde dann wieder von allen Teammitgliedern durchgeführt.

Mitnehmen kann ich aus Human-Computer-Interaction und dem damit verbundenem Projekt, dass die Einbeziehung der späteren Enduser viel zur Usability beiträgt. Durch die Erstellung von Prototypen und anschließende Usability-Interviews wurde man auf viele Usability-Probleme aufmerksam und konnte diese dann auch verbessern. Auch hat man durch die Durchführung Aspekte die die Usability verbessern und Methoden mit denen die Usability mit Endusern oder auch ohne getestet werden kann, kennen gelernt.

Insgesamt bin ich mit dem Endprodukt unseres Projektes sehr zufrieden. Was mir jedoch weniger gefallen hat, ist die Dokumentation der Ergebnisse und der Prototypen,

da die Ausformulierungen oft viel Zeit in Anspruch genommen haben. Auch ist es für mich schwierig genug Personen für die Interviews bzw. für das Experiment zu finden, wobei ich froh war, dass dies dann unser sozialerer Kollege größtenteils übernommen hat. Interessant fand ich jedoch wiederum die Ergebnisse aus dem Usability-Feedback und aus den Experimenten, da man hierdurch sieht, wie der High-Fidelity-Prototyp von den Usern gesehen wird.

6.2.2 Philipp Hiermann

Als sehr lehrreich empfand ich die Konzepte der Low- bzw. High-Fidelity-Prototypen. Weniger sinnvoll fand ich die Befragungen bei Meilenstein 4. Während es bei Meilenstein 3 für das Projekt durchaus hilfreich war, User-Feedback einzuholen, halte ich es für weniger sinnvoll, aus der Befragung von 10 Personen Schlüsse zu ziehen. Schließlich geht es bei dieser Lehrveranstaltung um die Mensch-Computer-Interaktion. Ich zweifle nicht daran, dass die Methode der Befragung und die anschließende statistische Auswertung der Antworten ein wichtiger Bestandteil des gesamten Entwicklungsprozesses ist, allerdings ist es zweifelhaft, aus der Befragung von 10 Personen genug Informationen zu extrahieren, um ein Ändern/Beibehalten des Designs zu rechtfertigen. Die Stichprobe ist einfach zu klein. Mir persönlich brachte der Meilenstein 4 also keinen Mehrwert.

6.2.3 Pascal Attwenger

Ich persönlich sehe den praktischen (Projekt-)Teil der LV HCI als einen kleinen, recht rudimentären Einblick in den Berufsalltag eines HCI-Spezialisten. Dass zum Entwickeln einer benutzerfreundlichen Softwarelösung mehr dazugehört als nur der reine, technische Entwicklungsprozess, sollte eigentlich einleuchten. Beim Durcharbeiten der Meilensteine ist mir dennoch aufgefallen, dass all die unterstützenden Aufgaben – Problemanalyse, Prototyping, Evaluierung etc. – doch einen deutlich größeren Stellenwert einnehmen als ich vermutet hätte.

Aufgrund dieser Vermutung ist wohl auch zu erklären, dass ich mir zu Beginn des Semesters von der Lehrveranstaltung etwas tiefergehendere Einblicke in das tatsächliche UI- bzw. UX-Design erwartet hätte. Zwar wurden Aspekte wie die Gestaltpsychologie oder Farbwahrnehmung im Vorlesungsteil durchaus besprochen, bei der Projektumsetzung traten sie jedoch wieder stark in den Hintergrund. Insgesamt schien mir, dass das Design des HiFi-Prototypen bei den meisten Gruppen – uns selbst nicht ausgenommen – mehr auf einem Trial-and-Error-Prinzip basierte als auf den theoretischen Grundlagen.

Beim Vorlesungsteil selbst erging es mir ähnlich: Hier hatte ich großes Interesse an den Einheiten mit psychologischem Hintergrund, während ich fand, dass die Teile zu den „handwerklichen“ Bereichen des HCI-Metiers versuchten, Praxiserfahrungen zu simulieren, was m.M.n. meist nicht so gut gelang.

Es sollte wohl im gesamten Studium mehr auf die Thematik eingegangen werden; wenn „gutes“ Design auch in anderen LVs ein Thema wäre, dann könnte hier umso besser auf die Arbeit damit eingegangen werden, ohne dabei Details zu verlieren.