

# Statistical Analysis using IBM SPSS Statistics Version 23

## Questionnaire Items

Dependent variables:

The questionnaire contained the following questions:

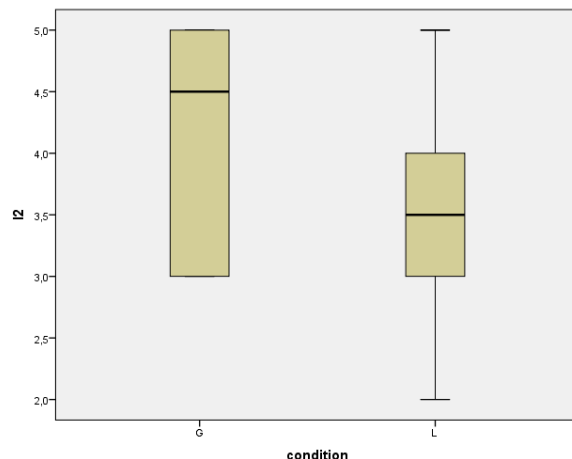
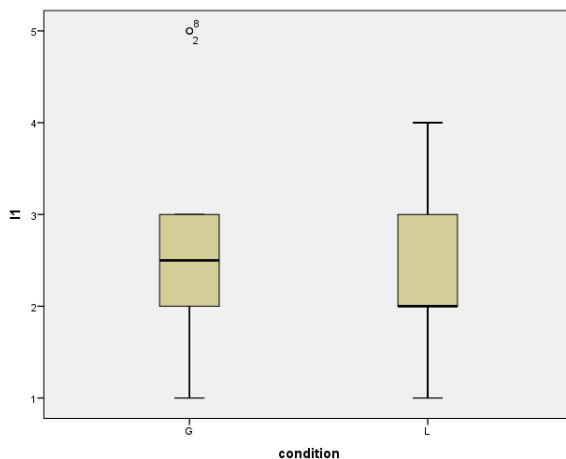
Please answer the questions below with a value from **1** (= I fully **disagree**) to **5** (= I fully **agree**):

- I1. I had a very good **overview of the documents** I had already visited.
- I2. It was very easy to **find and revisit key documents again**
- I3. It was very easy to **find the relevant passages** in the key documents again.
- D1. It was very easy to **find open documents** on the large display
- D2. It was very easy to **manage (place, re-size...) multiple windows** on the large display
- S1. It was very easy to **search** within the provided documents.
- O1. I am very satisfied with the **outcome** of this session
- O2. The **large display hardware** was very pleasant to use.
- O3. Overall, the **software** was very pleasant to use

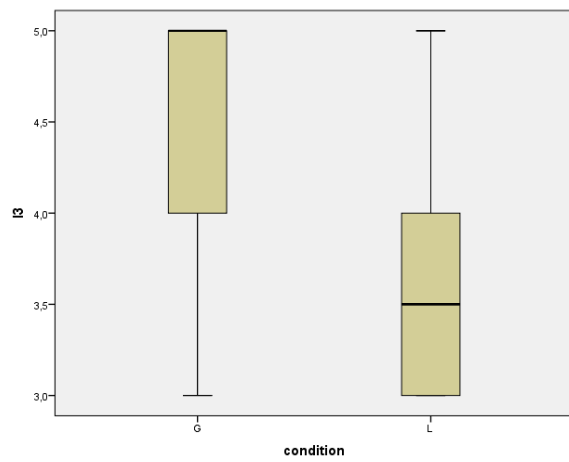
Please rate how much the items listed below helped you to solve the task with a value from **1** (= not at all) to **5** (= very much).

- (H1. Concept Graph)<sup>1</sup>
- H2. Search Tool
- H3. Large Display Space
- H4. Visual Links

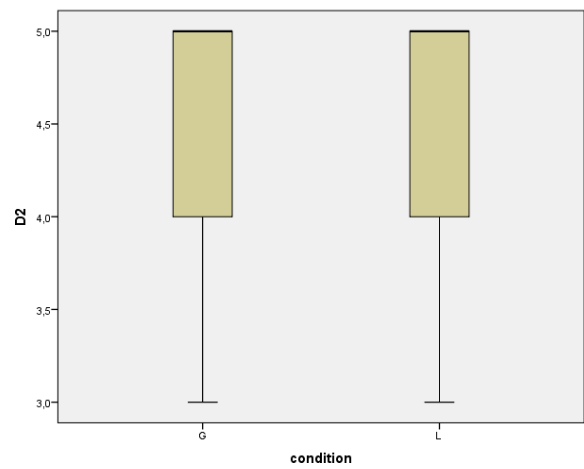
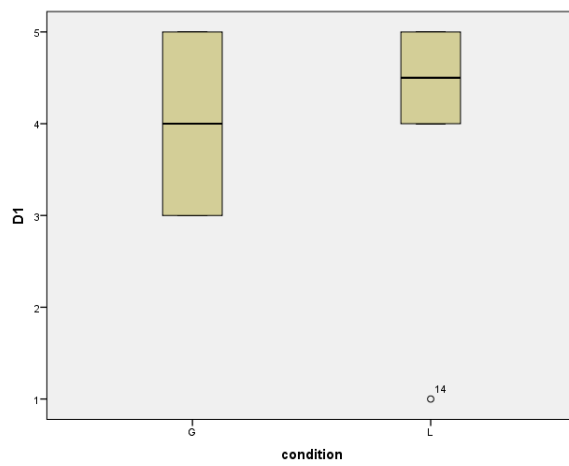
Outlier test:



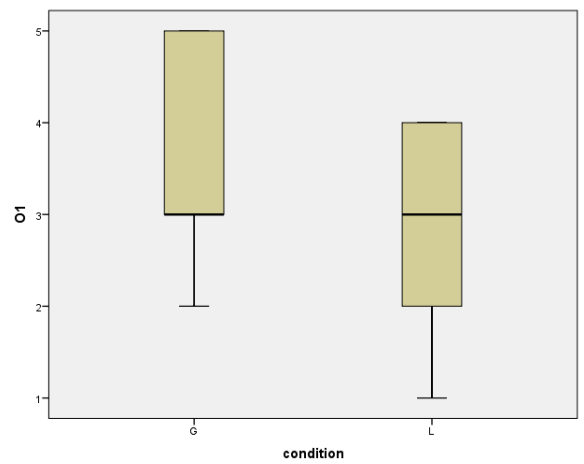
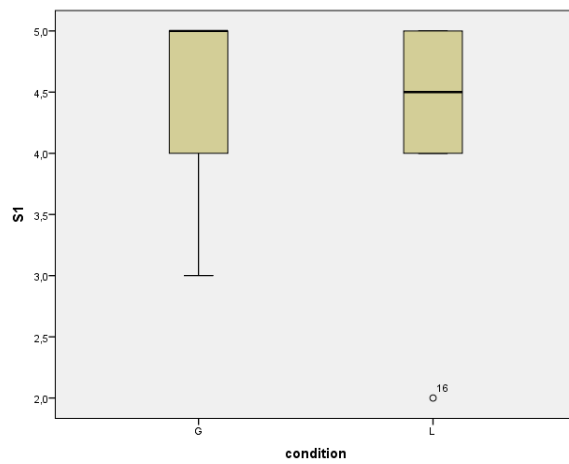
<sup>1</sup> Not evaluated here, because it was only rated by one group.

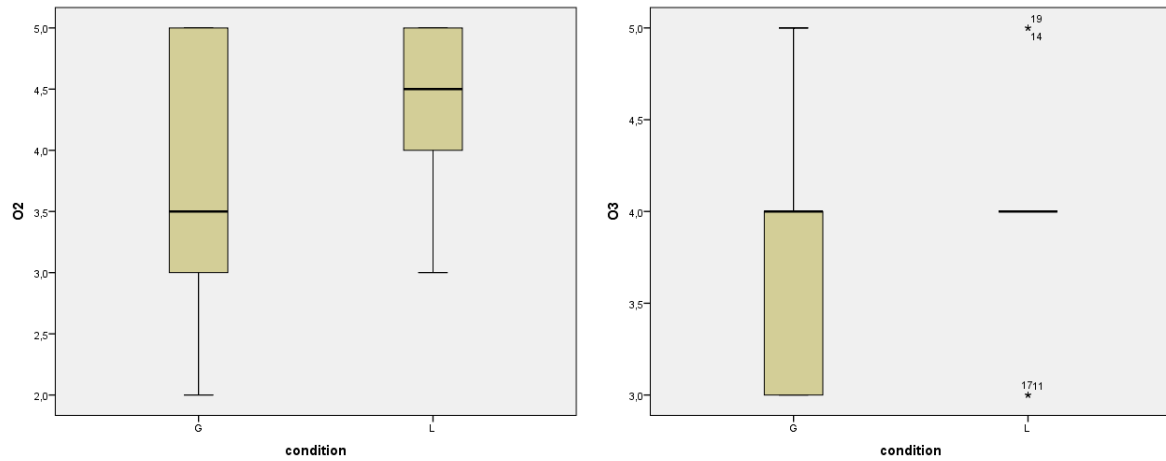


For I1, two outliers were detected (PG2 and PG8).

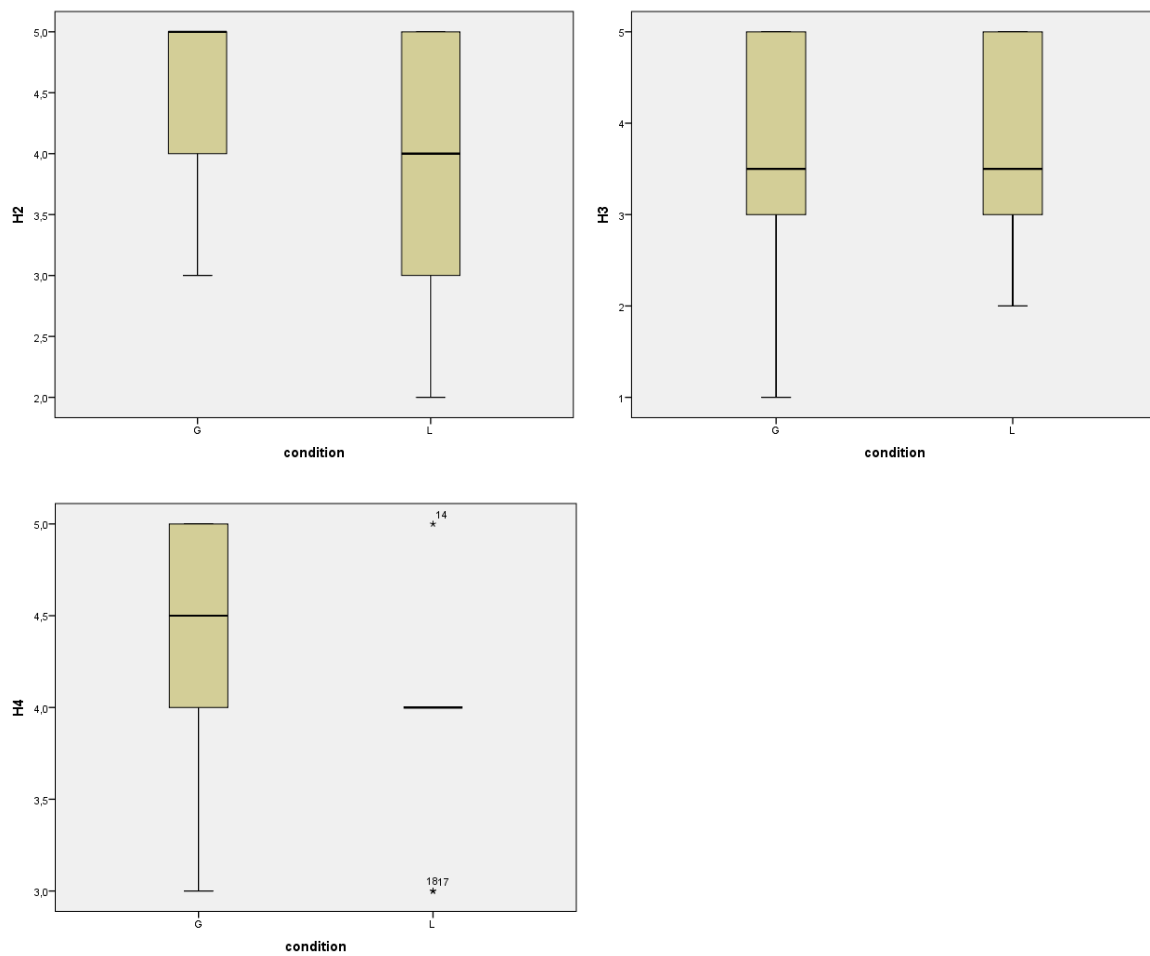


No outliers detected for D1 and D2.





We detected an outlier (PL6) for S1. For O3, we detected four outliers (PL1, PL4, PL7, PL9).



For H4, we found 3 outliers.

All outliers were manually removed before performing any further tests.

Normality tests:

Tests auf Normalverteilung							
condition		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
I1	G	,240	10	,108	,858	10	,073
	L	,245	10	,090	,892	10	,177
I2	G	,308	10	,008	,756	10	,004
	L	,233	10	,133	,904	10	,245
I3	G	,360	10	,001	,731	10	,002
	L	,302	10	,010	,781	10	,008

a. Signifikanzkorrektur nach Lilliefors

For I2 and I3, the normality assumption is violated ( $p < .05$ ).

Tests auf Normalverteilung							
condition		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
D1	G	,256	10	,063	,769	10	,006
	L	,335	10	,002	,662	10	,000
D2	G	,360	10	,001	,731	10	,002
	L	,416	10	,000	,650	10	,000

a. Signifikanzkorrektur nach Lilliefors

For D1 and D2, the normality assumption is also violated.

Tests auf Normalverteilung							
condition		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
S1	G	,360	10	,001	,731	10	,002
	L	,276	10	,030	,727	10	,002
O1	G	,333	10	,002	,793	10	,012
	L	,229	10	,148	,859	10	,074
O2	G	,246	10	,089	,874	10	,111
	L	,302	10	,010	,781	10	,008
O3	G	,254	10	,067	,833	10	,036
	L	,300	10	,011	,815	10	,022

a. Signifikanzkorrektur nach Lilliefors

For S1, O1, O2, and O3 the normality assumption is also violated.

Tests auf Normalverteilung							
condition		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
H2	G	,422	10	,000	,628	10	,000
	L	,240	10	,107	,886	10	,152
H3	G	,234	10	,127	,835	10	,039
	L	,195	10	,200*	,878	10	,124
H4	G	,305	10	,009	,781	10	,008
	L	,370	10	,000	,752	10	,004

\*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

For H2 and H4, the normality assumption is violated. We therefore performed independent t-tests for I1 and H3, and Mann-Whitney U tests for the remaining questionnaire items.

## Independent t-tests

Gruppenstatistiken

condition	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
H3 G	10	3,70	1,337	,423
L	10	3,60	1,174	,371
I1 G	8	2,25	,707	,250
L	10	2,40	1,075	,340

The table above shows the descriptive statistics (mean and standard deviation).

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit					
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz
									Untere      Obere
H3	Varianzen sind gleich	,141	,712	,178	18	,861	,100	,563	-1,082      1,282
	Varianzen sind nicht gleich			,178	17,702	,861	,100	,563	-1,084      1,284
I1	Varianzen sind gleich	1,978	,179	-,339	16	,739	-,150	,442	-1,087      ,787
	Varianzen sind nicht gleich			-,355	15,528	,727	-,150	,422	-1,047      ,747

Since the Levene Tests for Equality of Variances are not statistically significant (red box), we can assume equal group variances. We did not find any significant differences for H3 and I1.

## Mann-Whitney U Test

Ränge

conditionNum	N	Mittlerer Rang	Rangsumme
I2 1,00	10	12,70	127,00
2,00	10	8,30	83,00
Gesamt	20		
I3 1,00	10	13,10	131,00
2,00	10	7,90	79,00
Gesamt	20		
D1 1,00	10	8,50	85,00
2,00	9	11,67	105,00
Gesamt	19		
D2 1,00	10	10,05	100,50
2,00	10	10,95	109,50
Gesamt	20		
S1 1,00	10	10,00	100,00
2,00	9	10,00	90,00
Gesamt	19		
O1 1,00	10	11,15	111,50
2,00	10	9,85	98,50
Gesamt	20		
O2 1,00	10	8,85	88,50
2,00	10	12,15	121,50
Gesamt	20		
O3 1,00	10	8,20	82,00
2,00	6	9,00	54,00
Gesamt	16		
H2 1,00	10	12,35	123,50
2,00	10	8,65	86,50
Gesamt	20		
H4 1,00	10	10,40	104,00
2,00	7	7,00	49,00
Gesamt	17		

Condition 1 corresponds to the BLC condition, and condition 2 stands for the links condition.

Statistik für Test<sup>a</sup>

	I2	I3	D1	D2	S1	O1	O2	O3	H2	H4
Mann-Whitney-U	28,000	24,000	30,000	45,500	45,000	43,500	33,500	27,000	31,500	21,000
Wilcoxon-W	83,000	79,000	85,000	100,500	90,000	98,500	88,500	82,000	86,500	49,000
Z	-1,734	-2,090	-1,324	-,404	,000	-,513	-1,318	-,398	-1,514	-1,626
Asymptotische Signifikanz (2-seitig)	,083	,037	,185	,687	1,000	,608	,187	,691	,130	,104
Exakte Signifikanz [2*(1- seitige Sig.)]	,105 <sup>b</sup>	,052 <sup>b</sup>	,243 <sup>b</sup>	,739 <sup>b</sup>	1,000 <sup>b</sup>	,631 <sup>b</sup>	,218 <sup>b</sup>	,792 <sup>b</sup>	,165 <sup>b</sup>	,193 <sup>b</sup>

a. Gruppenvariable: conditionNum

b. Nicht für Bindungen korrigiert.

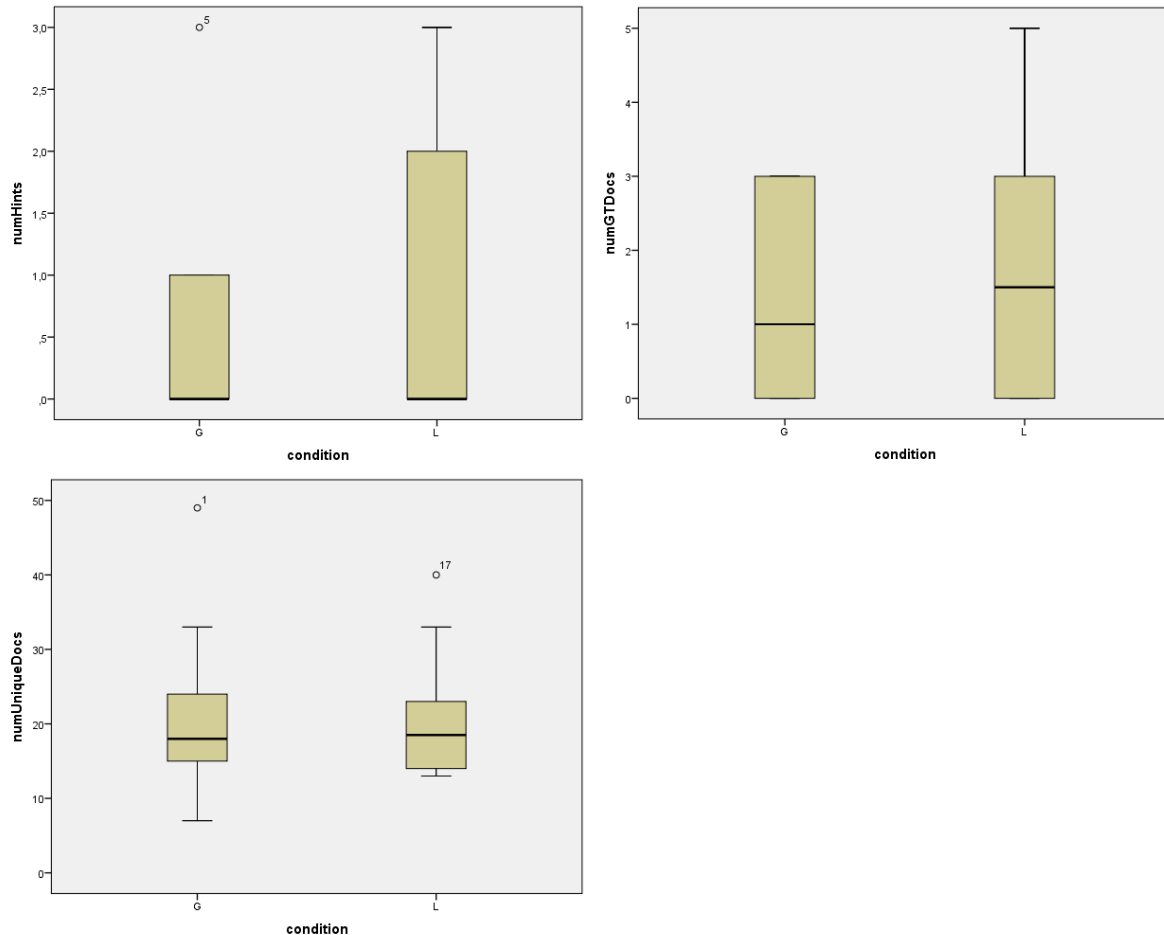
The only significant difference was found for I3 ( $p=.037$ ).

## Task Performance

### Dependent variables:

- numHints: number of correctly identified hints
- numGTDocs: number of distinct documents opened out of the 13 ground truth documents
- numUniqueDocs: number of distinct documents opened (both, distractors and ground truth)

### Outlier tests:



Outliers found for numHints and numUniqueDocs were manually removed before performing any further tests.

### Normality tests:

Tests auf Normalverteilung						
condition	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
numHints G	,397	10	,000	,603	10	,000
numHints L	,359	10	,001	,740	10	,003
numGTDocs G	,234	10	,127	,803	10	,016
numGTDocs L	,225	10	,164	,867	10	,092
numUniqueDocs G	,217	10	,199	,880	10	,132
numUniqueDocs L	,224	10	,169	,846	10	,052

a. Signifikanzkorrektur nach Lilliefors

The normality assumption is violated for numHints. We therefore performed a Mann-Whitney U test for numHints, and independent t-tests for numGTDocs and numUniqueDocs.

## Independent t-tests

Test bei unabhängigen Stichproben									
		Levene-Test der Varianzhomogenität		T-Test für die Mittelwertgleichheit					
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz Untere      Obere
numGTDocs	Varianzen sind gleich	,549	,468	-,437	18	,668	-,300	,687	-1,744      1,144
	Varianzen sind nicht gleich			-,437	17,001	,668	-,300	,687	-1,750      1,150
numUniqueDocs	Varianzen sind gleich	,114	,740	-,239	16	,814	-,778	3,248	-7,664      6,109
	Varianzen sind nicht gleich			-,239	15,554	,814	-,778	3,248	-7,680      6,125

Since the Levene Tests for Equality of Variances are not statistically significant (red box), we can assume equal group variances. There is no significant difference for both dependent variables (green boxes).

## Mann-Whitney U test

Ränge				
	conditionNum	N	Mittlerer Rang	Rangsumme
numHints	1,00	9	8,78	79,00
	2,00	10	11,10	111,00
Gesamt		19		

The links condition has a higher mean rank for the number of correctly identified hints.

Statistik für Test <sup>a</sup>		numHints
Mann-Whitney-U		34,000
Wilcoxon-W		79,000
Z		-1,092
Asymptotische Signifikanz (2-seitig)		,275
Exakte Signifikanz [2*(1-seitige Sig.)]		,400 <sup>b</sup>

a. Gruppenvariable: conditionNum

b. Nicht für Bindungen korrigiert.

There is no significant difference between the two groups.

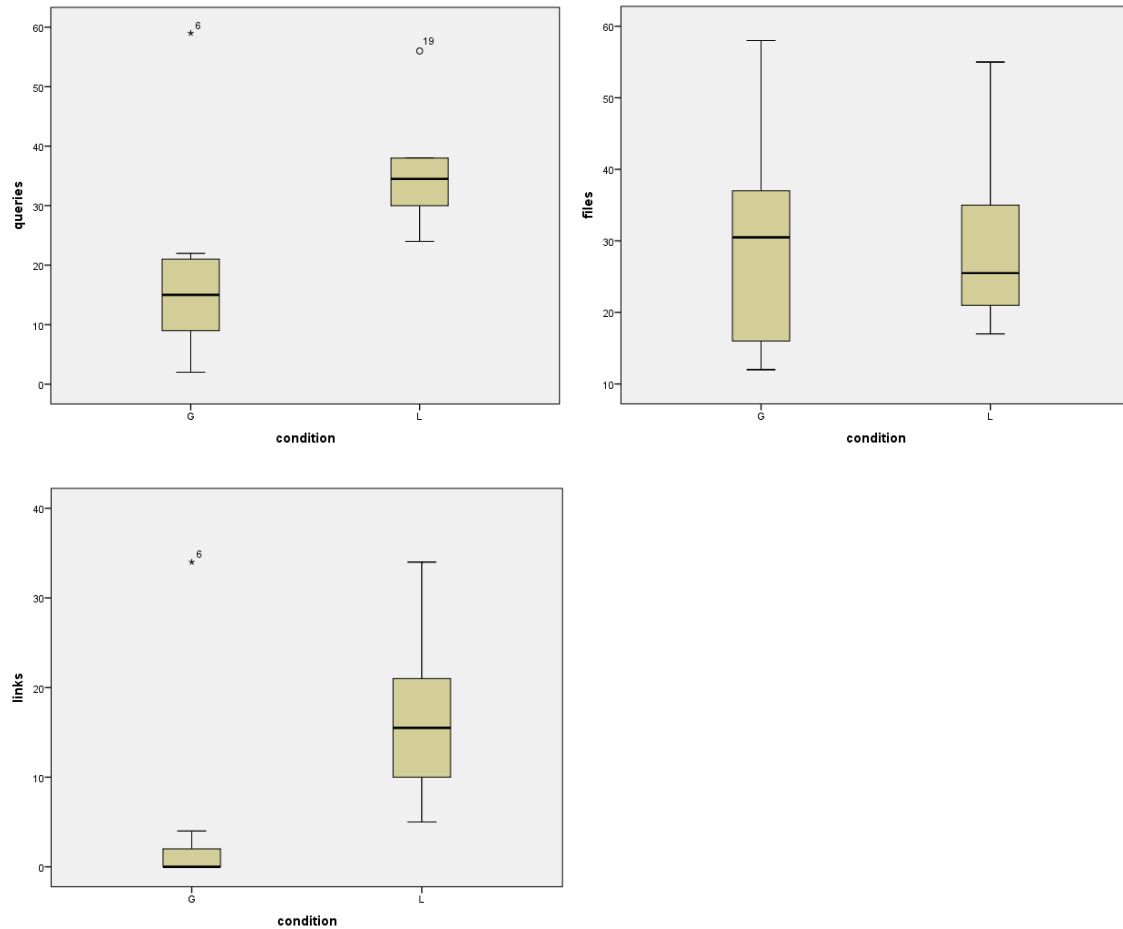


## Usage logs

### Dependent variables:

- Queries: number of queries performed in the search tool over the entire study
- Files: number of text files opened from the search tool over the entire study:
- Links: number of visual links initiated over the entire study

### Outlier tests:



We found two outliers for queries and one outlier for links. All outliers were manually removed before performing any further tests.

Normality tests:

Tests auf Normalverteilung							
condition		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
queries	G	,304	10	,010	,744	10	,003
	L	,272	10	,034	,858	10	,071
files	G	,152	10	,200 <sup>*</sup>	,935	10	,500
	L	,305	10	,009	,843	10	,048
links	G	,404	10	,000	,453	10	,000
	L	,191	10	,200 <sup>*</sup>	,944	10	,603

\*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

The normality assumption is violated. We therefore performed Mann-Whitney U Tests for all three dependent variables.

Mann-Whitney U Test

Ränge				
conditionNum		N	Mittlerer Rang	Rangsumme
queries	1,00	9	5,00	45,00
	2,00	9	14,00	126,00
	Gesamt	18		
files	1,00	10	11,10	111,00
	2,00	10	9,90	99,00
	Gesamt	20		
links	1,00	9	5,00	45,00
	2,00	10	14,50	145,00
	Gesamt	19		

The links condition has a higher mean rank for the number of queries and visual links, but a lower mean rank for the number of opened files.

Statistik für Test <sup>a</sup>			
	queries	files	links
Mann-Whitney-U	,000	44,000	,000
Wilcoxon-W	45,000	99,000	45,000
Z	-3,580	-,454	-3,732
Asymptotische Signifikanz (2-seitig)	,000	,650	,000
Exakte Signifikanz [2*(1-seitige Sig.)]	,000 <sup>b</sup>	,684 <sup>b</sup>	,000 <sup>b</sup>

a. Gruppenvariable: conditionNum

b. Nicht für Bindungen korrigiert.

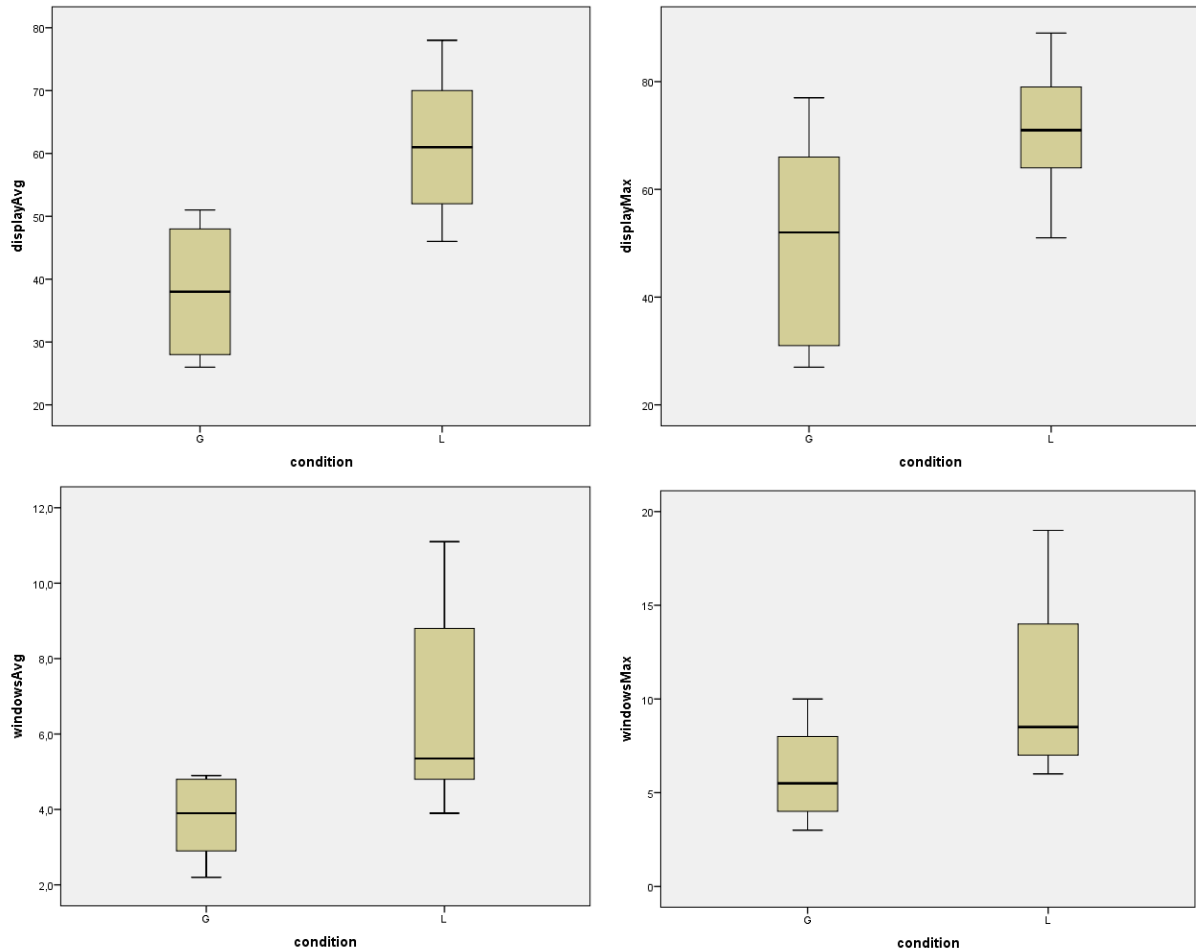
We found significant differences for the number of queries and visual links.

## Display Space and Window Usage

### Dependent variables:

- DisplayAvg: Average usage of display space over the entire study (in % of overall display space)
- displayMax: Maximum usage of display space during the study (in % of overall display space)
- windowsAvg: Average number of open windows over the entire study
- windowsMax: Maximum number of open windows during the study

### Outlier tests:



We found no outliers for the dependent variables.

## Normality test:

Tests auf Normalverteilung

condition		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
displayAvg	G	,188	10	,200 <sup>*</sup>	,868	10	,096
	L	,122	10	,200 <sup>*</sup>	,972	10	,906
displayMax	G	,187	10	,200 <sup>*</sup>	,914	10	,309
	L	,137	10	,200 <sup>*</sup>	,981	10	,971
windowsAvg	G	,164	10	,200 <sup>*</sup>	,926	10	,412
	L	,281	10	,024	,838	10	,042
windowsMax	G	,150	10	,200 <sup>*</sup>	,947	10	,630
	L	,228	10	1,50	,859	10	,075

\*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

All dependent variables, except for windowsAvg follow a normal distribution ( $p > .05$ ). We can therefore conduct independent sample t-tests for displayAvg, displayMax, and windowsMax, and a Mann-Whitney U Test for windowsAvg.

## Independent t-tests

Gruppenstatistiken

condition		N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
displayAvg	G	10	38,20	10,261	3,245
	L	10	60,80	9,987	3,158
displayMax	G	10	49,20	17,862	5,649
	L	10	70,90	11,210	3,545
windowsMax	G	10	5,90	2,331	,737
	L	10	10,30	4,270	1,350

The table above shows the descriptive statistics (mean and standard deviation).

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
displayAvg	Varianzen sind gleich	,528	,477	-4,991	18	,000	-22,600	4,528	-32,113	-13,087
	Varianzen sind nicht gleich			-4,991	17,987	,000	-22,600	4,528	-32,113	-13,087
displayMax	Varianzen sind gleich	2,895	,106	-3,254	18	,004	-21,700	6,669	-35,711	-7,689
	Varianzen sind nicht gleich			-3,254	15,137	,005	-21,700	6,669	-35,903	-7,497
windowsMax	Varianzen sind gleich	3,469	,079	-2,860	18	,010	-4,400	1,538	-7,632	-1,168
	Varianzen sind nicht gleich			-2,860	13,926	,013	-4,400	1,538	-7,701	-1,099

Since the Levene Tests for Equality of Variances are not statistically significant (red box), we can assume equal group variances. All three independent variables are statistically significant (green boxes).

## Mann-Whitney U Test

Ränge				
	conditionNum	N	Mittlerer Rang	Rangsumme
windowsAvg	1,00	10	6,70	67,00
	2,00	10	14,30	143,00
	Gesamt	20		

Condition 2 (links) has a higher mean rank.

Statistik für Test <sup>a</sup>		windowsAvg
Mann-Whitney-U		12,000
Wilcoxon-W		67,000
Z		-2,884
Asymptotische Signifikanz (2-seitig)		,004
Exakte Signifikanz [2*(1-seitige Sig.)]		,003 <sup>b</sup>

a. Gruppenvariable: conditionNum

b. Nicht für Bindungen korrigiert.

This difference is statistically significant.