Data Analysis: "Manipulating Task Determinability for studying alternative arrangements of User Interface"

RQ1: What is the effect of task determinability on user interaction with experimental interface compared to baseline interface?

Hypothesis 1a:

If a task is specified (S) task, study proposes null hypothesis. No major differences will be observed in users interacting with experimental interface and baseline interface because pre-determinable tasks are easy to achieve, and users will be able to complete such tasks easily on both the interfaces in similar manner.

Data Analysis 1a:

- Descriptive Statistics:

Interaction Variable	Number of Participants	Mean (Average)	Standard Deviation
Number of Documents	18	9.83	2.01
Number of Queries	18	2.5	0.985
Time taken in (minutes)	18	5.27	0.761

Table 1 Mean and Standard deviation of User performing Specified Task on Klink Interface

Interaction Variable	Number of Participants	Mean (Average)	Standard Deviation
Number of Documents	18	8.83	3.11
Number of Queries	18	2.83	0.707
Time taken in (minutes)	18	5.32	1.22

Table 2 Mean and Standard deviation of User performing Specified Task on IEEE Interface

- Box plots showing five point summary

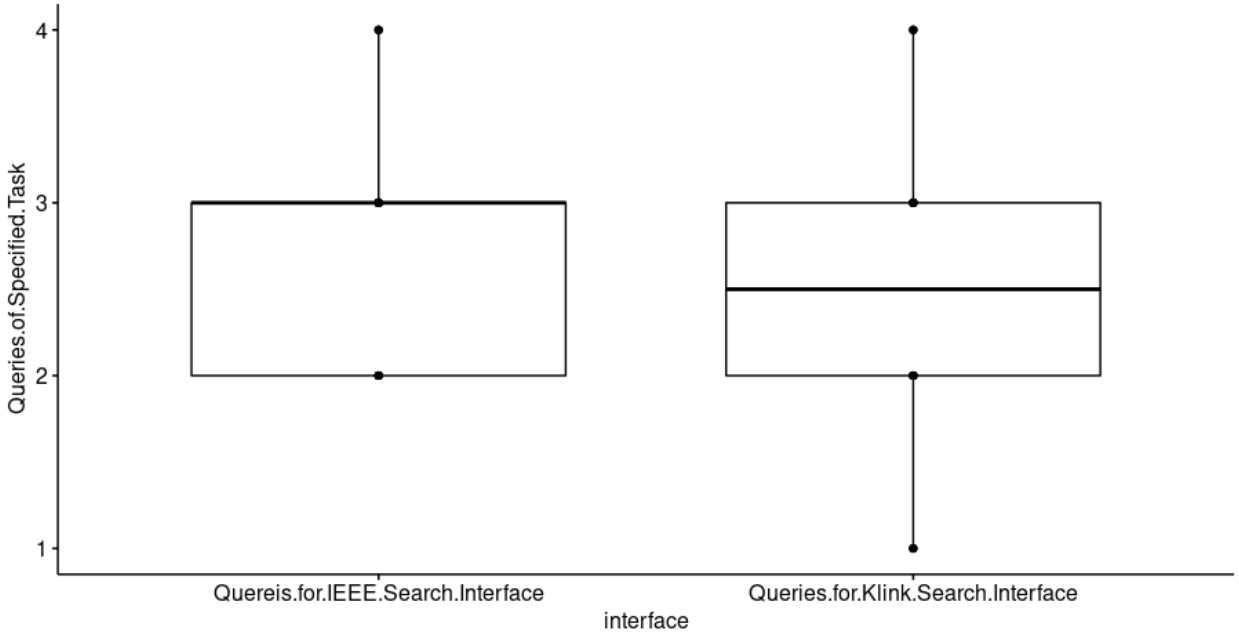


Fig 1 Box plots showing five point summary of queries required on both interfaces completing a specific task

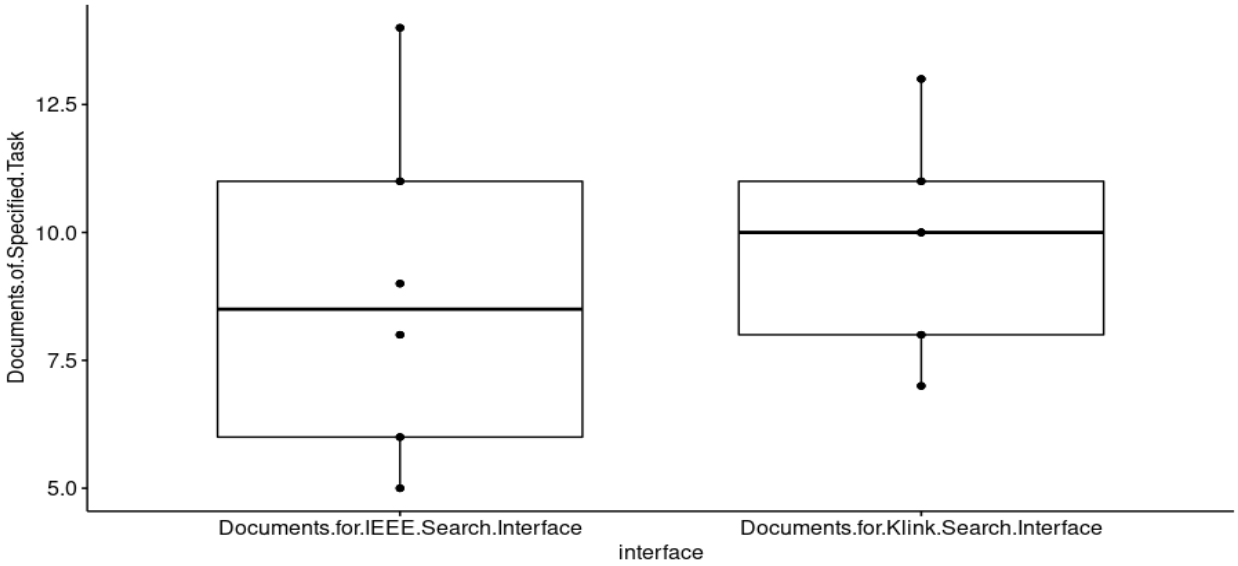


Fig 2 Box plots showing five point summary for documents saved on both interfaces completing a specific task

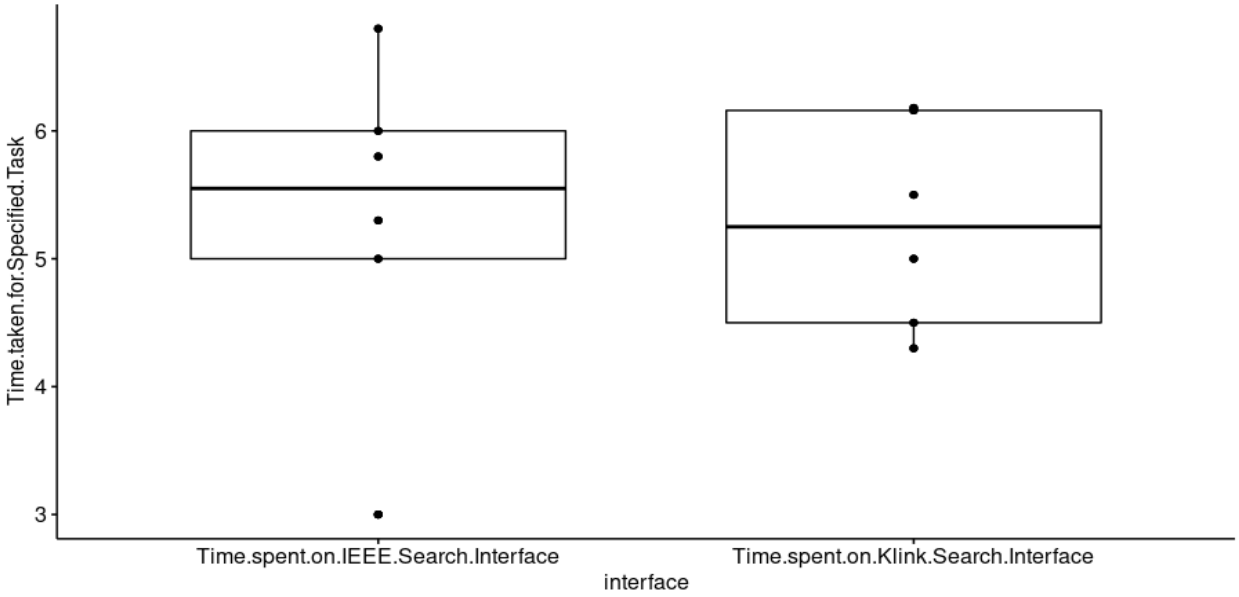


Fig 3 Box plots showing five point summary for time spent on both interfaces completing a specific task

- Inferential Statistics 1a:

1. Hypothesis testing for Number of documents saved completing a specified task on both the interfaces

   Test for Normality (Shapiro-Wilk test) Results:

	Statistic	P-Value
Documents for IEEE Search Interface (Specified Task)	0.901	0.0604
Documents for Klink Search Interface (Specified Task)	0.901	0.0603

Since p>0.05 we can accept null hypothesis and hence our distribution is normal

Hypothesis Testing (Repeated Measures Anova for within subject study design)

F (1,17) = 1.041	P=0.32 > 0.05	Generalized effect size= 0.037

Since p value is greater than 0.05 we accept null hypothesis, which underlines that there is no significant difference between data from both the interfaces

2. Hypothesis testing for Number of queries required completing a specified task on both the interfaces

   Test for Normality (Shapiro-Wilk test) Results:

	Statistic	P-Value
Queries required for IEEE Search Interface (Specified Task)	0.807	0.00191
Queries required for Klink Search Interface (Specified Task)	0.893	0.0428

Since p < 0.05 we can accept alternative hypothesis and hence our distribution is not normal

Hypothesis Testing (Wilcoxon signed rank test for within subject study design)

V=0	Z-stat -2.332847	p= 0.01966 < 0.05	We can reject null hypothesis and accept alternative hypothesis

3. Hypothesis testing for total time spent completing a specified task on both the interfaces

   Test for Normality (Shapiro-Wilk test) Results:

	Statistic	P-Value
Time Spent for IEEE Search Interface (Specified Task)	0.849	0.00829
Time Spent for Klink Search Interface (Specified Task)	0.845	0.00706

   Since the value of p in both the cases is less than 0.05 we accept alternative hypothesis and distribution of data is not normal

   Hypothesis Testing (Wilcoxon signed rank test for within subject study design)

V=51	Z-stat -1.489746	p= 0.1363 > 0.05	We can accept null hypothesis and reject alternative hypothesis

Hypothesis 1b: If a task is unspecified (U) task, study proposes alternative hypothesis i.e. User will interact more with baseline interface than experimental interface. Since experimental interface supports evaluation of search results with additional features like (list of keywords), users will need to spend less effort and time to complete open-ended tasks on experimental interface than baseline interface. So, users will have to spend more effort, might have to traverse multiple pages and hence must interact more with baseline interface in order to find relevant information

- Descriptive Statistics:

Interaction Variable	Number of Participants	Mean (Average)	Standard Deviation

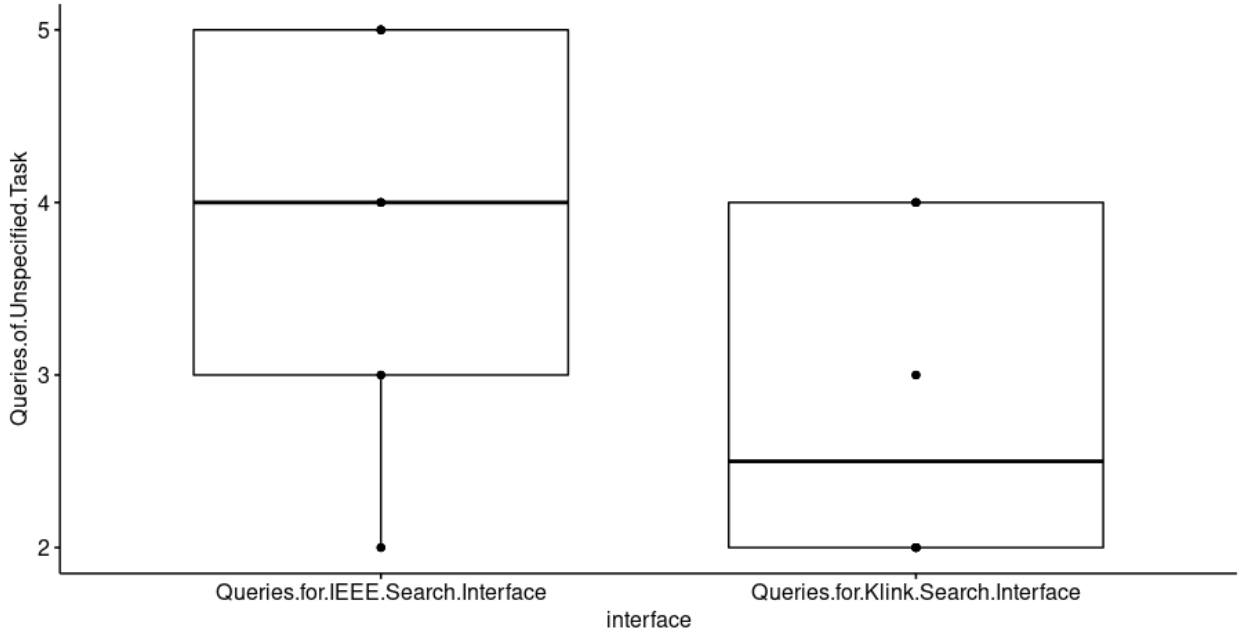
Number of Documents	18	10	1.57
Number of Queries	18	2.83	0.92
Time taken in (minutes)	18	5.2	0.696

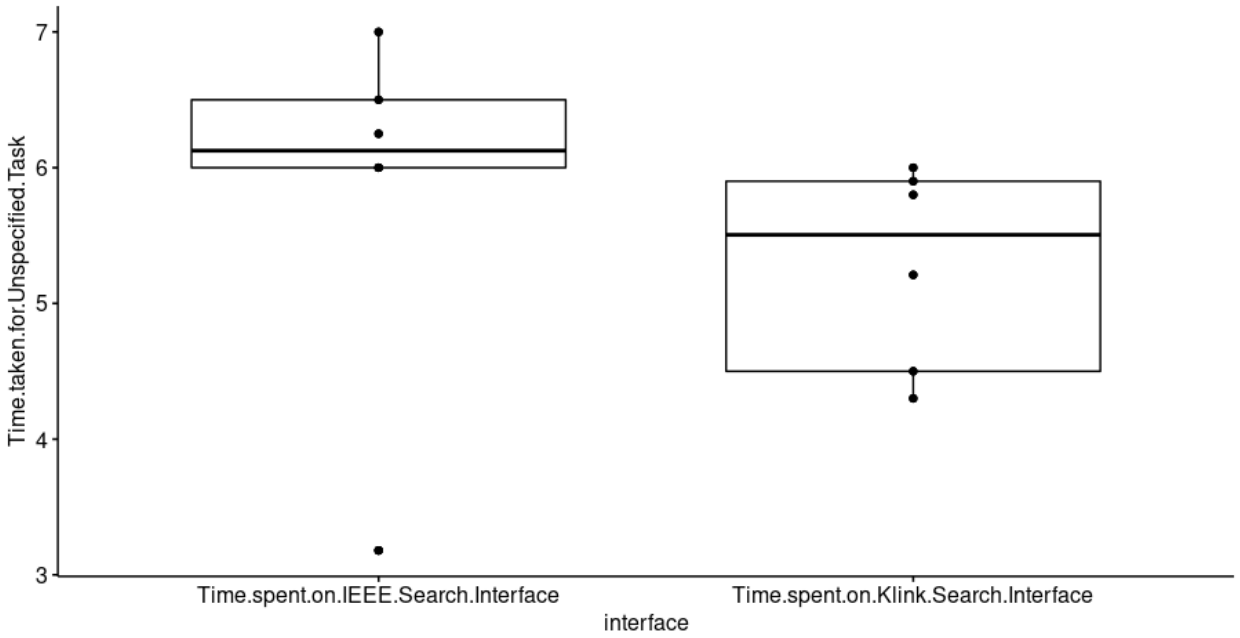
Table 3: Mean and Standard deviation of User performing Unspecified Task on Klink Interface

Interaction Variable	Number of Participants	Mean (Average)	Standard Deviation
Number of Documents	18	8.33	1.94
Number of Queries	18	3.83	1.10
Time taken in (minutes)	18	5.8	1.26

Table 4: Mean and Standard deviation of User performing Unspecified Task on IEEE Interface

- Box Plots for five point summary





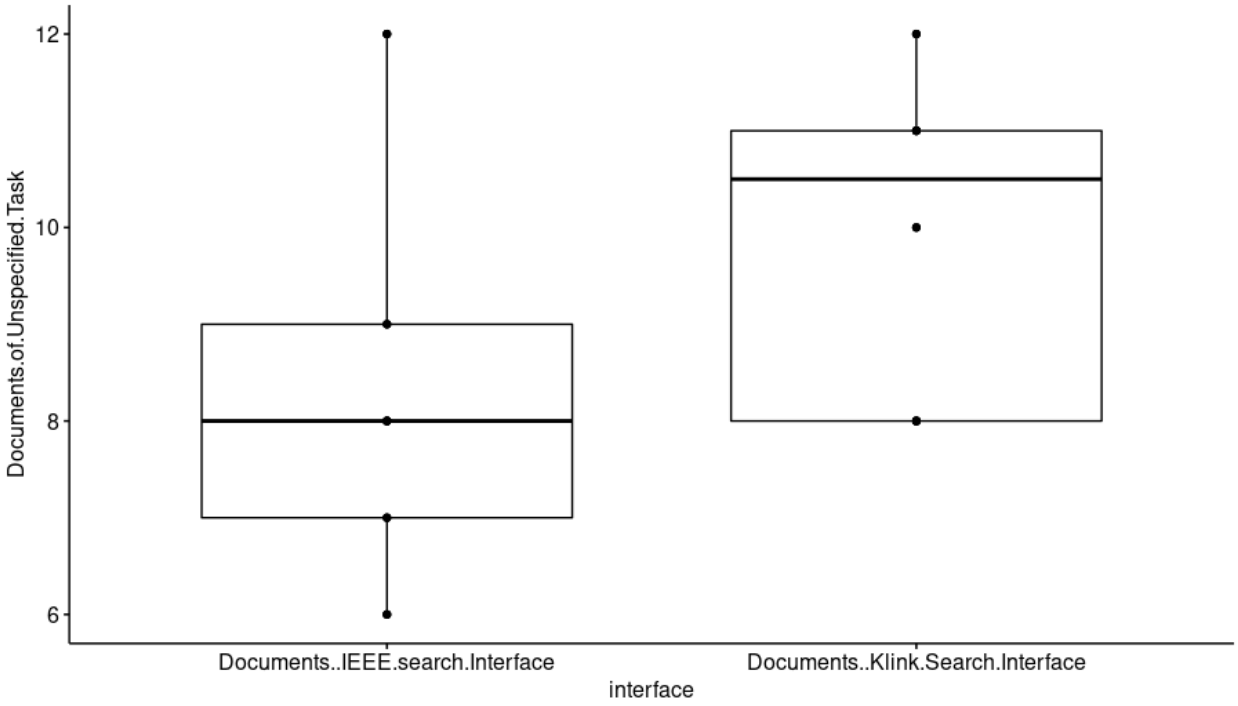


Fig 4 Box Plots for Queries, documents, time taken for completing specified task on both the interfaces

- Inferential Statistics:

Normality test (Shapiro-Wilk test) Results:

	Statistic	P-Value	Decision About Normality
Queries required for IEEE Search Interface (Unspecified Task)	0.848	0.00785	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Queries required for Klink Search Interface	0.729	0.000175	p-value < 0.05 Accept Alternative

(Unspecified Task)			Hypothesis, hence data is not normal
Time required for IEEE Search Interface (Unspecified Task)	0.699	0.0000770	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Time required for Klink Search Interface (Unspecified Task)	0.807	0.00193	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Documents saved for IEEE Search Interface (Unspecified Task)	0.850	0.00847	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Documents saved for Klink Search Interface (Unspecified Task)	0.820	0.00297	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal

Table 5 Normality test for User interaction factors while Unspecified Task

Hypothesis Testing (Wilcoxon signed rank test for within subject study design)

Interaction factor (Unspecified Task)	V-value	Z stat	P-value	Accepted / Rejected Null Hypothesis

Documents saved on Klink and IEEE interface	V = 129,	-1.92832	= 0.05381	Probably Type I error but if you consider results from repeated measures Anova then we can accept alternative hypothesis
Queries required on Klink and IEEE interface	V = 24,	-2.07040	= 0.03841	Rejected null hypothesis because of statistical significance. There is a difference in data from both the interfaces
Time spent on Klink and IEEE interface	V = 51,	-1.48494	= 0.1376	Accept null hypothesis because there is no evidence to accept alternative hypothesis

Table 6 Hypothesis Testing for User interaction (Unspecified Task)

RQ2: What is the effect of task determinability on user experience with experimental interface compared to baseline interface?

Hypothesis 2a: If a task is specified (S) task, study proposes null hypothesis. Since these tasks are easy to achieve and hence independent of interface design, users will not much invest their cognitive and emotional effort in completing the search tasks. Therefore, study hypothesize that users will not experience significant differences while completing information specified (S) search task on experimental interface

compared to baseline interface. Hence, user rating of each sub-scale of user experience for experimental interface will be identical to baseline interface.

- Descriptive Statistics:

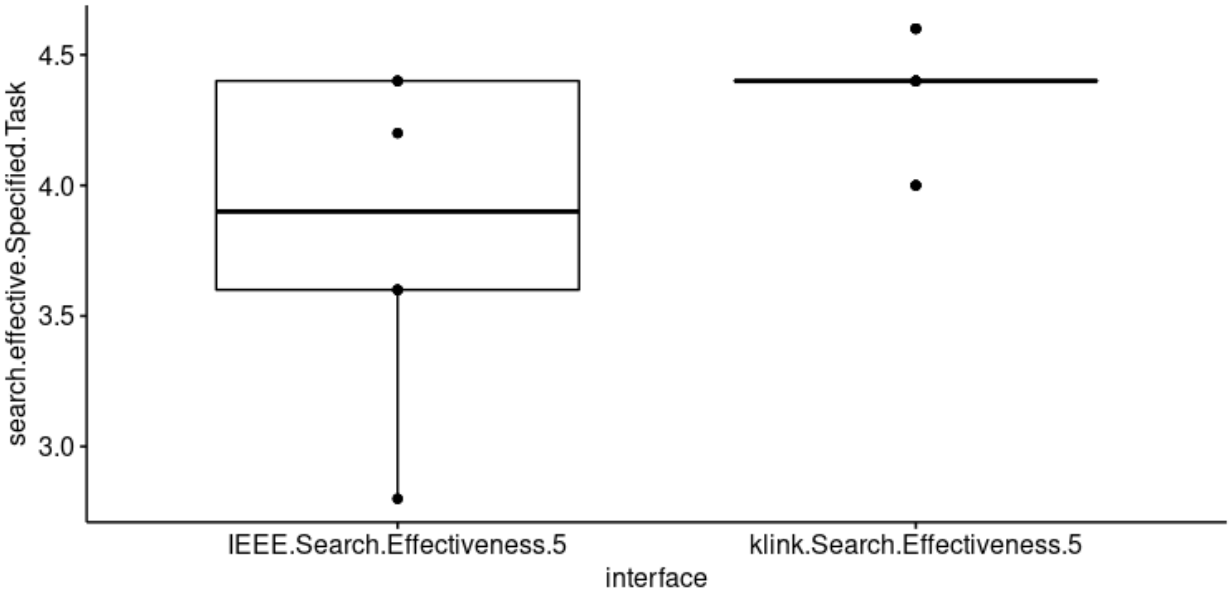
Sub factor for User Engagement Scale (Specified Task on Klink Search Interface)	Number of Participants	Mean (Average)	Standard Deviation
Focused Attention	18	3.72	1.04
Perceived Usability	18	4.72	0.366
Search Effectiveness	18	4.37	0.185
Reward	18	4.61	0.502

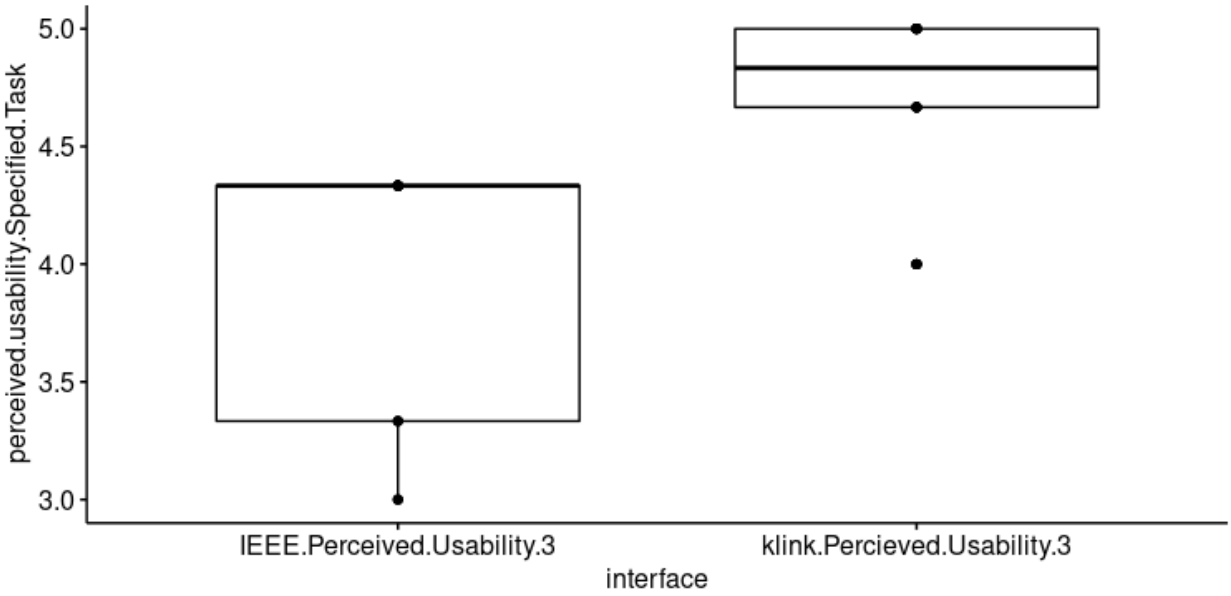
Table 7 Mean and Standard deviation for user engagement sub-scale factor (Klink Interface)

Sub factor for User Engagement Scale (Specified Task on IEEE Search Interface)	Number of Participants	Mean (Average)	Standard Deviation
Focused Attention	18	3.78	0.856
Perceived Usability	18	3.94	0.575
Search Effectiveness	18	3.83	0.587
Reward	18	3.83	0.328

Table 8 Mean and Standard deviation for user engagement sub-scale factor (IEEE Interface)

- Box Plots for five-point summary





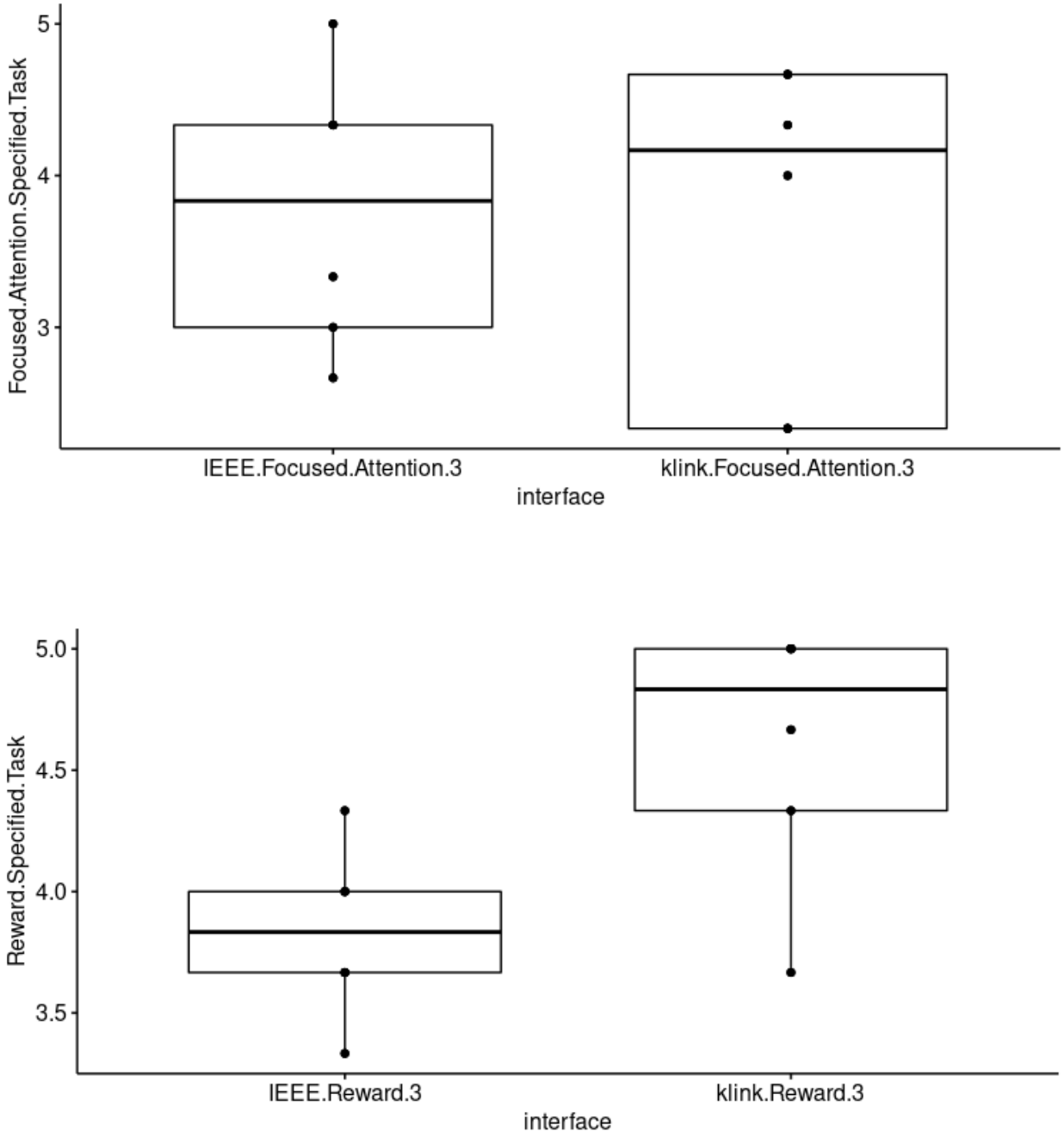


Fig 5 Box plots for User engagement Factors (Specified Task)

- **Inferential Statistics**:

Normality test (Shapiro-Wilk test) Results:

User engagement factor for Specified Task	Statistics	p-value	Decision about Normality
Focused Attention for Klink Interface	0.736	0.000214	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Focused Attention for IEEE Interface	0.868	0.0162	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Perceived Usability for Klink Interface	0.718	0.000132	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Perceived Usability for IEEE Interface	0.648	0.0000209	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Search Effectiveness for Klink Interface	0.709	0.000102	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal

Search Effectiveness for IEEE Interface	0.815	0.00248	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Reward for Klink Search Interface	0.754	0.000365	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Reward for IEEE Search Interface	0.893	0.0428	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal

Table 9 Normality Test for User engagement factors for Specified Task

Hypothesis Testing (Wilcoxon signed rank test for within subject study design)

User engagement Factor	V-Value	Z-stat	p-value	Hypothesis Accepted / Rejected
Focused Attention	V = 76.5,	-0.37364	= 0.7087	P > 0.05. Hence Accept null hypothesis
Perceived Usability	V = 165,	-3.46146	= 0.0005372	P < 0.05 Accept alternative Hypothesis
Search effectiveness	V = 72,	-2.56534	= 0.01031	P < 0.05 Accept

				alternative Hypothesis
Reward	V = 156,	-3.05723	= 0.002234	P < 0.05 Accept alternative Hypothesis

Table 10 Hypothesis Testing for User Engagement factors (Specified Task)

Hypothesis 2b: If a task is unspecified (U) task, study proposes alternative hypothesis i.e., user will rate each sub scale of user experience higher for experimental interface compared to baseline interface. As experimental interface supports features by which open-ended tasks can be completed effectively, user will rate better experience with experimental interface compared to baseline interface.

- Descriptive Statistics

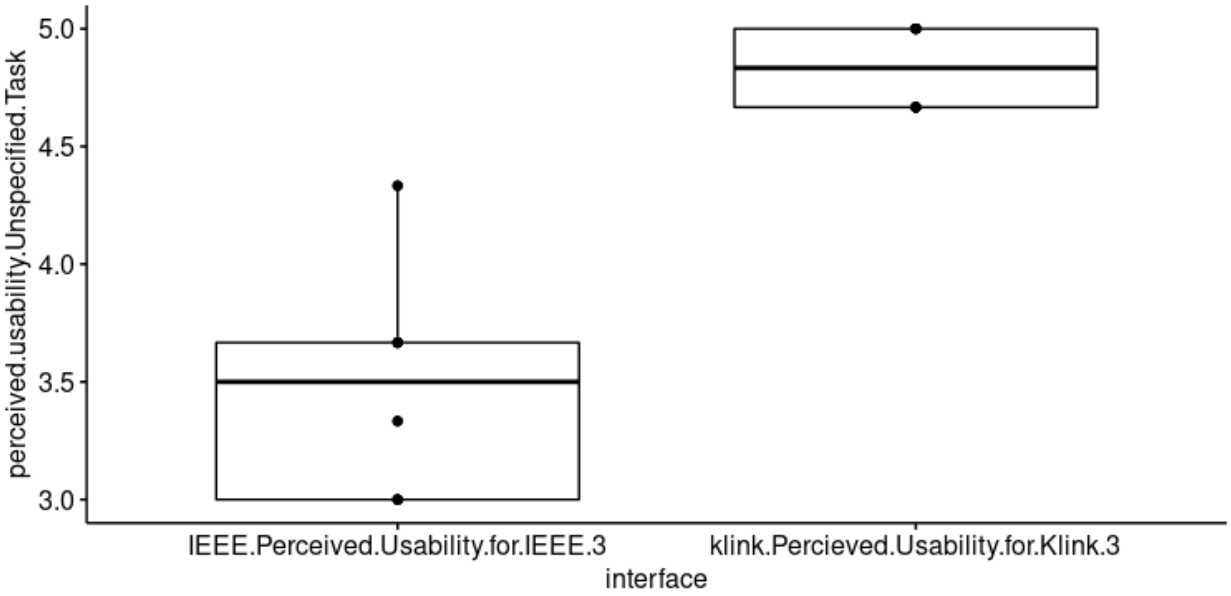
Sub factor for User Engagement Scale (Unspecified Task on Klink Search Interface)	Number of Participants	Mean (Average)	Standard Deviation
Focused Attention	18	4.56	0.256
Perceived Usability	18	4.83	0.171
Search Effectiveness	18	4.63	0.22
Reward	18	4.33	0.28

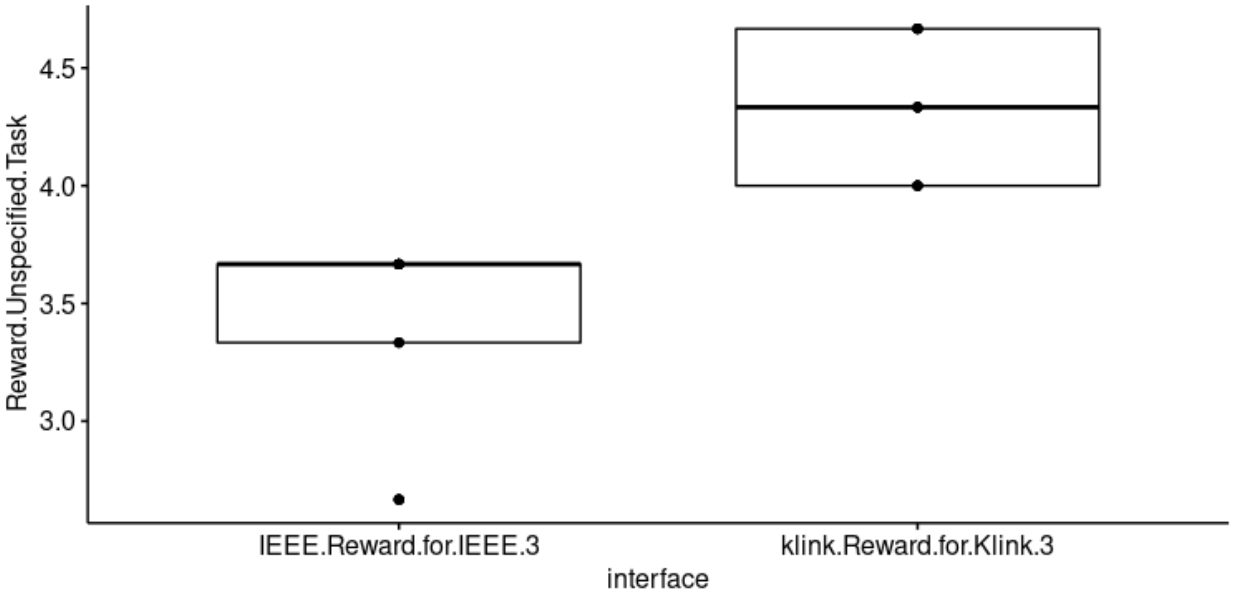
Table 11 Mean and Standard deviation for User Engagement factors (Klink Interface)

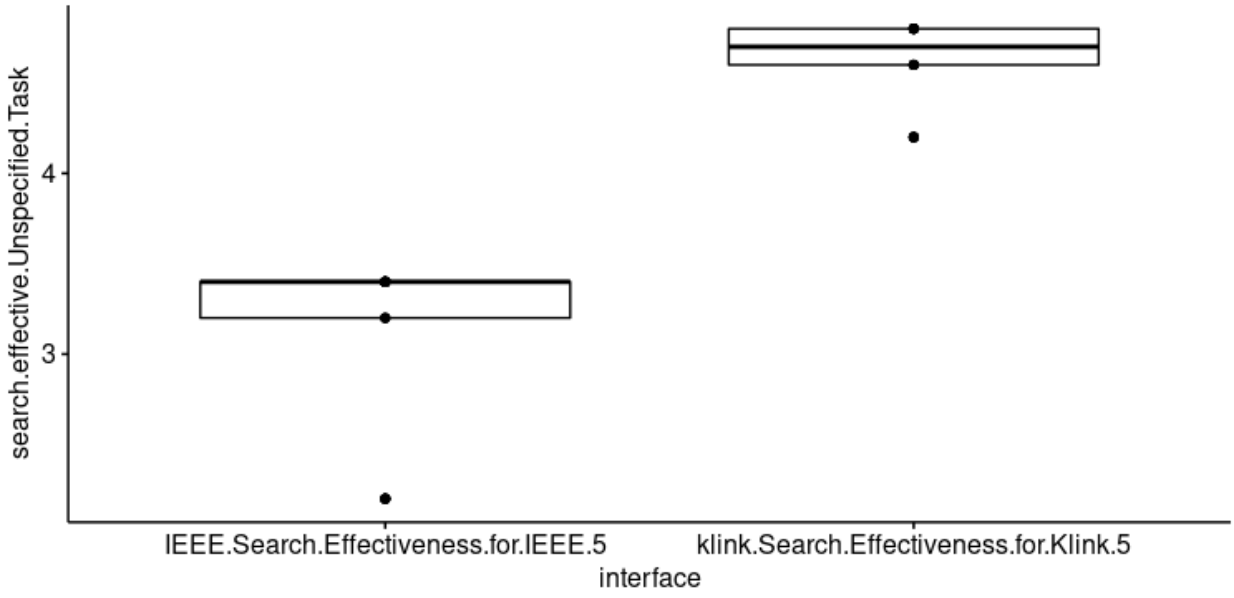
Sub factor for User Engagement Scale (Unpecified Task on IEEE Search Interface)	Number of Participants	Mean (Average)	Standard Deviation
Focused Attention	18	2.94	0.826
Perceived Usability	18	3.5	0.475
Search Effectiveness	18	3.17	0.451
Reward	18	3.44	0.379

Table 12 Mean and Standard deviation for User Engagement factors (IEEE Interface)

- Box plots for five points summaries




Reward.Unspecified.Task
4.5
4.0
3.5
3.0
IEEE.Reward.for.IEEE.3
klink.Reward.for.Klink.3
interface


search.effective.Unspecified.Task
4
3
IEEE.Search.Effectiveness.for.IEEE.5
klink.Search.Effectiveness.for.Klink.5
interface

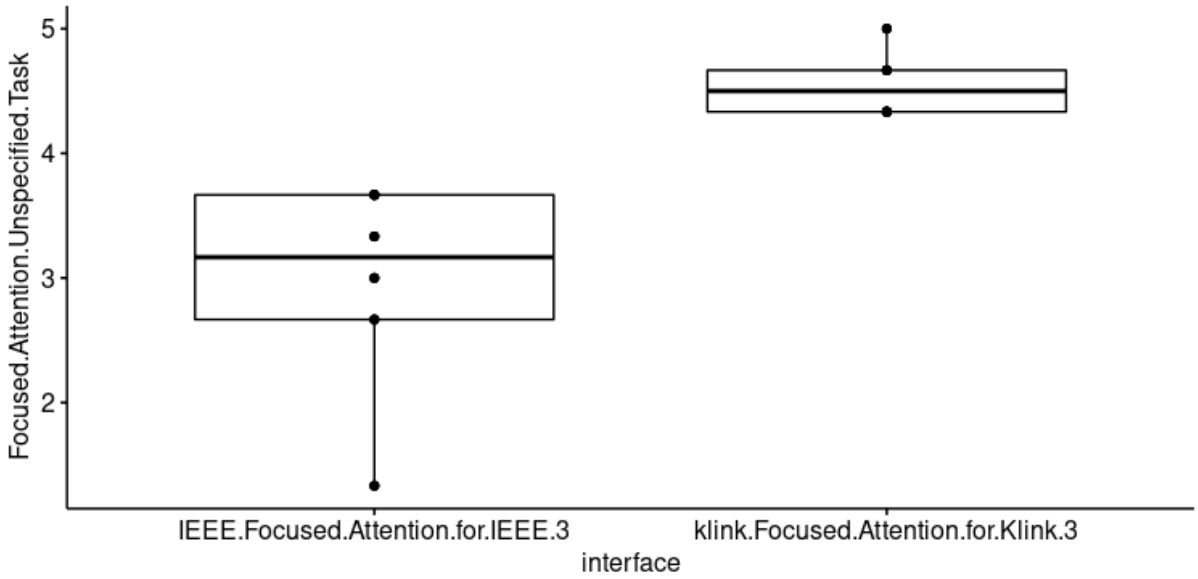


Fig 6 Box plots for User engagement factors (Unspecified Task)

- Inferential Statistics

Normality test (Shapiro-Wilk test) Results:

User engagement factor for Unspecified Task	Statistics	p-value	Decision about Normality
Focused Attention for Klink Interface	0.767	0.000534	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Focused Attention for IEEE Interface	0.788	0.00104	p-value < 0.05 Accept Alternative Hypothesis,

			hence data is not normal
Perceived Usability for Klink Interface	0.642	0.0000183	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Perceived Usability for IEEE Interface	0.841	0.00623	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Search Effectiveness for Klink Interface	0.718	0.000132	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Search Effectiveness for IEEE Interface	0.549	0.00000222	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Reward for Klink Search Interface	0.801	0.00157	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal
Reward for IEEE Search Interface	0.618	0.0000103	p-value < 0.05 Accept Alternative Hypothesis, hence data is not normal

Table 13 Normality test User engagement Factors (Unspecified Task)

Hypothesis Testing (Wilcoxon signed rank test for within subject study design)

User engagement Factor	V-Value	Z-stat	p-value	Hypothesis Accepted / Rejected
Focused Attention	V = 171,	-3.73648	= 0.0001866	P < 0.05. Hence Accept Alternative hypothesis
Perceived Usability	V = 171,	-3.73648	= 0.0001866	P < 0.05 Accept alternative Hypothesis
Search effectiveness	V = 171,	-3.72436	= 0.0001958	P < 0.05 Accept alternative Hypothesis
Reward	V = 171,	-3.73648	= 0.0001866	P < 0.05 Accept alternative Hypothesis

Table 14 Hypothesis Testing User engagement factors (Unspecified Task)

Conclusion

The main purpose of this study was to manipulate task determinability to evaluate alternative arrangements of user interface, hence we can evaluate how much an interface arrangement can assist users to perform differing levels of search task activities. Thus, we have measured variables like user interaction and user engagement by studying participants on both the interfaces. Experimental results show that,

- If the task is unspecified, then user engagement levels differ on both the interfaces, thus we accept alternative hypothesis because of p values less than 0.05. Overall engagement score for Klink interface is higher than IEEE interface. Same results are obtained in user performing specified task for user engagement on both the interfaces. Only exception in this is case is sub-factor "Focused Attention" where p=0.70 and we have rejected alternative hypothesis.
- If the task is unspecified, then there is significant difference observed in number of queries required to complete unspecified task on both the interfaces. There is no significant difference in total time spent to complete unspecified task on both the interfaces. For number of documents saved for Unspecified task, p=0.053 which might be an indicator of Type I error.
- If the task is specified, there is a significant difference observed in number of queries required to complete task on both the interfaces. For documents and time spent there were no statistically significant result achieved.