

# Participants' Data Collected From the Context and Debugging User Studies

# List of Tables

1.1	Context Study: Results of each subject's success in completing each task . . . . .	2
1.2	Context Study: Results of the subjects' time on each task (in minutes). . . . .	3
1.3	Context Study: Lostness results for the tasks for 18 subjects. . . . .	4
1.4	Context Study: Result of the subjects' errors on each task. . . . .	5
1.5	Context Study: Post-session experience ratings. . . . .	6
1.6	Context Study: Pre-task expectation ratings. . . . .	7
1.7	Context Study: Post-task experience ratings. . . . .	8
1.8	Summary results of verbal comments made during the Context study. . . . .	9
1.9	Statistics related to analysed metrics in the Context user study. . . . .	9
2.1	Debugging Study: Results of each subject's success in completing each task. . . . .	11
2.2	Debugging Study: Result of the subjects' errors on each task. . . . .	12
2.3	Debugging Study: Results of the subjects' time on each task (in minutes). . . . .	13
2.4	Debugging Study: Lostness results for the tasks for 18 subjects. . . . .	14
2.5	Debugging Study: Post-session experience ratings. . . . .	15
2.6	Debugging Study: Pre-task expectation ratings. . . . .	16
2.7	Debugging Study: Post-task experience ratings. . . . .	17
2.8	Debugging Study: Results of the disruption-related questions. . . . .	18
2.9	Summary results of verbal comments made during the Debugging study (part 1). . . . .	18
2.10	Summary results of verbal comments made during the Debugging study (part 2). . . . .	18
2.11	Statistics related to analysed metrics in the Debugging user study. . . . .	19



# Chapter 1

## Results of the Context User Study

Table 1.1: Context Study: Results of each subject's success in completing each task

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8	Task9
P1	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
P2	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	0.0
P3	0.5	1.0	0.5	0.5	1.0	0.0	0.5	0.0	0.0
P4	1.0	1.0	0.5	1.0	1.0	1.0	1.0	0.5	0.0
P5 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5
P6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P7 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P8 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5
P9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0
P10 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P11 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P12 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P13 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P14 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P15 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P16	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
P17	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5
P18	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0
mean	0.89	1.00	0.89	0.89	1.00	0.83	0.94	0.78	0.39
median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
stdev	0.22	0.00	0.22	0.22	0.00	0.35	0.17	0.36	0.49
mean*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89
median*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
stdev*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22

\* Denotes the subjects who used our tool.

Table 1.2: Context Study: Results of the subjects' time on each task (in minutes).

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8	Task9
P1	04:42	04:11	04:10	05:10	02:19	02:13	03:06	01:54	16:06
P2	08:43	07:39	03:14	02:07	00:44	02:16	02:57	02:36	18:40
P3	04:12	03:25	05:11	02:23	01:37	05:13	06:48	04:31	11:00
P4	06:40	02:59	06:21	01:57	00:23	04:07	02:50	08:41	12:48
P5 *	04:08	06:06	02:03	01:51	00:56	00:50	02:34	02:32	26:24
P6	02:59	02:47	03:27	01:43	01:44	02:19	02:46	02:22	15:09
P7 *	02:27	04:49	02:36	03:30	01:17	01:30	02:00	04:46	11:06
P8 *	02:41	02:40	03:02	02:21	00:57	01:47	02:38	02:05	11:45
P9	10:12	03:50	03:36	05:45	02:07	04:09	01:43	03:55	12:34
P10 *	02:54	01:55	05:08	02:52	00:50	01:58	03:19	02:34	13:30
P11 *	03:44	07:49	05:40	05:25	01:02	02:49	04:49	04:31	19:18
P12 *	02:14	04:55	02:24	02:09	00:47	01:27	02:29	02:32	09:41
P13 *	02:55	03:59	02:33	01:44	01:10	01:12	02:31	02:14	10:51
P14 *	01:55	03:08	01:25	01:38	00:45	00:49	02:30	01:57	09:46
P15 *	02:10	04:02	01:58	03:11	02:29	01:40	02:43	02:03	13:19
P16	09:37	07:06	06:21	04:21	01:36	05:36	04:48	04:19	14:20
P17	06:18	04:16	04:24	02:32	01:46	02:34	04:52	03:49	13:21
P18	08:04	04:23	04:46	02:34	01:53	03:11	04:50	02:56	17:36
geo-mean	06:22	04:16	04:29	02:54	01:24	03:18	03:34	03:32	14:26
median	06:40	04:06	03:32	02:27	01:13	02:15	02:48	02:35	13:20
stdev	02:31	01:43	01:32	01:20	00:36	01:25	01:19	01:40	04:10
geo-mean*	02:43	04:03	02:43	02:33	01:04	01:27	02:45	02:39	13:13
median*	02:41	04:02	02:33	02:21	00:57	01:30	02:34	02:32	11:45
stdev*	00:44	01:48	01:27	01:12	00:32	00:37	00:49	01:04	05:30

\* Denotes the subjects who used our tool.

Table 1.3: Context Study: Lostness results for the tasks for 18 subjects.

Participant	Task1		Task2		Task3		Task4		Task5		Task6		Task7		Task8		Task9	
	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L
P1	1,2,4	0.71	1,2,5	0.78	1,2,3	0.60	1,2,9	0.92	1,1,1	0.00	1,2,7	0.87	1,2,2	0.50	2,3,4	0.42	2,2,4	0.50
P2	1,2,11	0.96	1,3,3	0.67	1,2,3	0.60	1,2,3	0.60	1,2,3	0.60	1,3,6	0.83	1,2,3	0.60	2,3,3	0.33	2,6,33	1.00
P3	1,1,1	0.00	1,1,1	0.00	1,2,4	0.71	1,1,1	0.00	1,1,1	0.00	1,2,5	0.78	1,1,1	0.00	2,1,1	0.00	2,2,5	0.60
P4	1,2,2	0.50	1,3,3	0.67	1,2,2	0.50	1,2,7	0.87	1,3,3	0.67	1,3,10	0.97	1,2,5	0.78	2,2,3	0.33	2,4,12	0.83
P5 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,1,1	0.00	2,2,4	0.50
P6	1,2,3	0.60	1,2,5	0.78	1,2,3	0.60	1,2,3	0.60	1,2,3	0.60	1,2,3	0.60	1,2,3	0.60	2,3,4	0.42	2,4,8	0.71
P7 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,2,2	0.00	2,2,2	0.00
P8 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,1,1	0.00	2,2,2	0.00
P9	1,2,5	0.78	1,1,1	0.00	1,2,5	0.78	1,2,5	0.78	1,1,1	0.00	1,2,11	0.96	1,2,3	0.60	2,3,4	0.42	2,3,6	0.60
P10 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,2,2	0.00	2,2,2	0.00
P11 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,2,2	0.00	2,2,2	0.00
P12 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,2,2	0.00	2,2,2	0.00
P13 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,2,2	0.00	2,2,2	0.00
P14 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,2,2	0.00	2,2,2	0.00
P15 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	2,2,2	0.00	2,2,2	0.00
P16	1,2,5	0.78	1,2,7	0.87	1,2,9	0.92	1,2,3	0.60	1,2,3	0.60	1,2,7	0.87	1,2,5	0.78	2,3,3	0.33	2,1,1	0.00
P17	1,2,7	0.87	1,2,5	0.78	1,2,5	0.78	1,2,5	0.78	1,2,3	0.60	1,2,3	0.60	1,2,3	0.60	2,3,4	0.42	2,3,8	0.71
P18	1,2,5	0.78	1,2,5	0.78	1,1,1	0.00	1,2,5	0.78	1,1,1	0.00	1,2,2	0.50	1,2,3	0.60	2,3,4	0.42	2,3,12	0.82
mean		0.66		0.59		0.61		0.66		0.34		0.78		0.56		0.34		0.64
median		0.78		0.78		0.60		0.78		0.60		0.83		0.60		0.42		0.71
stdev		0.28		0.34		0.26		0.27		0.32		0.17		0.23		0.13		0.28
efficient		11.11%		22.22%		11.11%		11.11%		44.44%		0.00%		11.11%		100.00%		11.11%
(< 0.5)																		
inefficient		88.89%		77.78%		88.89%		88.89%		55.56%		100.00%		88.89%		0.00%		88.89%
(>= 0.5)																		
mean *		0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.11
median*		0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.00
stdev*		0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.22
efficient		100.00%		100.00%		100.00%		100.00%		100.00%		100.00%		100.00%		100.00%		77.78%
(< 0.5) *																		
inefficient		0.00%		0.00%		0.00%		0.00%		0.00%		0.00%		0.00%		0.00%		22.22%
(>= 0.5) *																		

\* Denotes the subjects who used our tool.

Table 1.4: Context Study: Result of the subjects' errors on each task.

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8	Task9
P1	1	1	0	1	0	0	0	0	0
P2	3	1	2	1	1	0	0	0	3
P3	1	3	2	2	1	2	0	0	0
P4	2	1	1	2	1	0	1	0	0
P5 *	0	0	0	0	0	0	0	0	0
P6	2	2	1	1	0	0	0	0	4
P7 *	0	0	0	0	0	0	0	0	0
P8 *	0	0	0	0	0	0	0	0	0
P9	2	1	1	0	1	0	0	0	1
P10 *	0	0	0	0	0	0	0	0	0
P11 *	0	0	0	0	0	0	0	0	0
P12 *	0	0	0	0	0	0	0	0	0
P13 *	0	0	0	0	0	0	0	0	0
P14 *	0	0	0	0	0	0	0	0	0
P15 *	0	0	0	0	0	0	0	0	0
P16	1	0	1	0	0	0	0	0	0
P17	3	2	0	0	0	0	1	0	0
P18	2	2	1	0	2	0	0	0	2
mean	1.89	1.44	1.00	0.78	0.67	0.22	0.22	0.00	1.11
median	2.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
stdev	0.78	0.88	0.71	0.83	0.71	0.67	0.44	0.00	1.54
mean*	0	0	0	0	0	0	0	0	0
median*	0	0	0	0	0	0	0	0	0
stdev*	0	0	0	0	0	0	0	0	0

\* Denotes the subjects who used our tool.

Table 1.5: Context Study: Post-session experience ratings.

Participant	Context (Q1)	Switch (Q2)	Relations (Q3)	Satisfaction (Q4)	Capability (UMUX.Q1)	Usability (UMUX.Q2)
P1	4	6	4	6	6	7
P2	4	6	6	6	5	5
P3	7	4	7	5	3	2
P4	5	2	0	2	4	2
P5 *	1	2	1	6	6	7
P6	6	7	6	2	2	2
P7 *	1	2	1	6	7	7
P8 *	2	1	3	5	5	5
P9	6	7	5	2	3	2
P10 *	1	2	1	6	6	6
P11 *	2	1	2	6	5	6
P12 *	2	3	2	6	6	6
P13 *	3	2	3	6	6	5
P14 *	1	1	1	7	6	6
P15 *	2	2	2	5	6	6
P16	3	4	4	3	3	3
P17	5	6	5	3	2	4
P18	6	7	7	2	2	2
mean	5.11	5.44	4.89	3.44	3.33	3.22
median	5.00	6.00	5.00	3.00	3.00	2.00
stdev	1.27	1.74	2.15	1.74	1.41	1.79
mean*	1.67	1.78	1.78	5.89	5.89	6.00
median*	2.00	2.00	2.00	6.00	6.00	6.00
stdev*	0.71	0.67	0.83	0.60	0.60	0.71

\* Denotes the subjects who used our tool.



Table 1.6: Context Study: Pre-task expectation ratings.

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8	Task9
P1	2	2	2	2	2	2	2	2	5
P2	3	4	3	3	2	2	2	2	6
P3	1	1	1	1	1	1	1	4	6
P4	1	2	2	1	1	1	4	2	5
P5 *	2	3	1	2	2	2	4	3	7
P6	3	3	3	3	2	2	3	3	4
P7 *	3	3	4	3	3	4	5	5	6
P8 *	3	3	3	3	3	4	4	4	5
P9	2	3	3	3	2	1	6	4	6
P10 *	2	3	4	4	2	3	5	6	7
P11 *	2	3	4	2	2	4	5	4	6
P12 *	3	5	5	2	2	2	5	3	6
P13 *	3	3	4	4	2	3	5	4	6
P14 *	3	4	3	3	2	2	4	3	6
P15 *	3	4	3	2	2	5	5	3	6
P16	2	2	3	3	2	2	3	4	6
P17	2	2	2	2	1	1	3	3	6
P18	1	1	2	2	1	1	2	2	5
mean	1.89	2.22	2.33	2.22	1.56	1.44	2.89	2.89	5.44
median	2.00	2.00	2.00	2.00	2.00	1.00	3.00	3.00	6.00
stdev	0.78	0.97	0.71	0.83	0.53	0.53	1.45	0.93	0.73
mean*	2.67	3.44	3.44	2.78	2.22	3.22	4.67	3.89	6.11
median*	3.00	3.00	4.00	3.00	2.00	3.00	5.00	4.00	6.00
stdev*	0.50	0.73	1.13	0.83	0.44	1.09	0.50	1.05	0.60

\* Denotes the subjects who used our tool.

Table 1.7: Context Study: Post-task experience ratings.

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8	Task9
P1	4	4	4	6	5	2	3	3	5
P2	4	4	4	5	4	4	4	4	7
P3	2	2	4	2	1	3	3	5	7
P4	1	2	2	2	1	2	3	5	3
P5 *	1	1	1	1	1	1	1	1	5
P6	2	3	4	3	2	5	4	5	7
P7 *	1	2	1	2	1	1	2	2	4
P8 *	1	2	3	2	1	2	3	1	4
P9	4	6	4	6	5	4	7	6	7
P10 *	1	2	4	3	1	1	3	2	6
P11 *	2	3	3	2	2	2	3	3	5
P12 *	1	3	2	1	1	1	2	2	4
P13 *	1	2	1	1	1	1	2	1	3
P14 *	1	2	1	1	1	1	1	1	3
P15 *	2	3	1	2	2	1	2	1	3
P16	3	3	4	4	3	4	5	5	7
P17	3	4	3	2	1	2	5	5	7
P18	3	3	4	3	2	3	4	3	7
mean	2.89	3.44	3.67	3.67	2.67	3.22	4.22	4.56	6.33
median	3.00	3.00	4.00	3.00	2.00	3.00	4.00	5.00	7.00
stdev	1.05	1.24	0.71	1.66	1.66	1.09	1.30	1.01	1.41
mean*	1.22	2.22	1.89	1.67	1.22	1.22	2.11	1.56	4.11
median*	1.00	2.00	1.00	2.00	1.00	1.00	2.00	1.00	4.00
stdev*	0.44	0.67	1.17	0.71	0.44	0.44	0.78	0.73	1.05

\* Denotes the subjects who used our tool.

Table 1.8: Summary results of verbal comments made during the Context study.

	Overall		Context		Reuse		Satisfaction		Syntax	
	Negative	Positive	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.
No. of Comments on UCAnDo-Models	14	27	4	3	0	6	5	17	5	1
No. of Comments on Other Tools	45	11	21	1	8	4	9	5	7	1
No. of Subjects who Commented on UCAnDoModels (Out of 9)	6	9	4	3	0	6	3	8	4	1
No. of Subjects who Commented on Other Tools (Out of 9)	9	6	9	1	4	4	4	3	4	1

Table 1.9: Statistics related to analysed metrics in the Context user study.

Collected Metric	Tool	Mean	Median	Min	Max	Std. Dev.	<i>p-value</i>	<i>r</i>
Task Success	Our Tool	0.99	1.00	0.94	1.00	0.02	<b>0.0032</b>	<b>0.70</b>
	Other Tools	0.85	0.89	0.44	1.00	0.17		
Task Errors	Our Tool	0.00	0.00	0.00	0.00	0.00	<b>0.0001</b>	<b>0.90</b>
	Other Tools	0.81	0.89	0.22	1.22	0.37		
Task Time	Our Tool	3.91	3.73	2.65	6.12	1.10	<b>0.0171</b>	<b>0.57</b>
	Other Tools	5.17	5.20	3.92	6.45	0.68		
Lostness	Our Tool	0.01	0.00	0.00	0.06	0.02	<b>0.0003</b>	<b>0.85</b>
	Other Tools	0.60	0.61	0.23	0.69	0.14		
Expectation vs. Experience Ratings	Our Tool	-1.69	-1.78	-2.33	-0.78	0.48	<b>0.0004</b>	<b>0.84</b>
	Other Tools	1.31	1.33	0.22	2.11	0.53		
UMUX	Our Tool	5.94	6.00	5.00	7.00	0.58	<b>0.0022</b>	<b>0.68</b>
	Other Tools	3.28	3.00	2.00	6.50	1.50		

The statistics related to Task Success, Errors on Tasks, Time on Tasks, Lostness, Users' Expectation versus Experience Ratings, and the UMUX (Usability Metric for User Experience) questionnaire for the Context study are presented in Table 1.9. As shown by the *p values* in the table (*p values*  $\leq 0.05$ ), our tool *significantly* outperforms other tools with respect to all the hypotheses in the Context study. We also calculate the effect size *r* for the Mann-Whitney U test. The *r* values suggest a *high practical significance* (*r*  $\geq 0.5$ ) between the subjects who used the UCAnDoModels editor and the subjects who used other editors.



# Chapter 2

## Results of the Debugging User Study

Table 2.1: Debugging Study: Results of each subject's success in completing each task.

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8
P1	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0
P2	1.0	0.5	1.0	0.5	1.0	1.0	0.0	1.0
P3	1.0	0.5	1.0	0.5	0.5	1.0	0.5	0.0
P4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P5 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P7 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P8 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P10 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P11 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P12 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P13 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P14 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P15 *	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
P16	1.0	0.5	1.0	0.5	1.0	1.0	1.0	1.0
P17	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0
P18	1.0	1.0	0.5	1.0	0.0	1.0	1.0	1.0
mean	1.00	0.83	0.94	0.61	0.83	1.00	0.83	0.89
median	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00
stdev	0.00	0.25	0.17	0.42	0.35	0.00	0.35	0.33
mean*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
median*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
stdev*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

\* Denotes the subjects who used our tool.

Table 2.2: Debugging Study: Result of the subjects' errors on each task.

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8
P1	0	0	0	2	0	0	0	2
P2	0	1	0	0	1	0	2	0
P3	0	0	0	0	0	0	1	2
P4	1	0	0	0	0	1	0	0
P5 *	0	0	0	0	0	0	0	0
P6	0	0	0	0	0	0	0	1
P7 *	0	0	0	0	0	0	0	0
P8 *	0	0	0	0	0	0	0	0
P9	0	0	0	0	0	0	0	0
P10 *	0	0	0	0	0	0	0	0
P11 *	0	0	0	0	0	0	0	0
P12 *	0	0	0	0	0	0	0	0
P13 *	0	0	0	0	0	0	0	0
P14 *	0	0	0	0	0	0	0	0
P15 *	0	0	0	0	0	0	0	0
P16	0	0	0	0	5	2	0	0
P17	1	0	0	2	0	0	0	0
P18	0	0	1	0	8	0	0	1
mean	0.22	0.11	0.11	0.44	1.56	0.33	0.33	0.67
median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
stdev	0.44	0.33	0.33	0.88	2.92	0.71	0.71	0.87
mean*	0	0	0	0	0	0	0	0
median*	0	0	0	0	0	0	0	0
stdev*	0	0	0	0	0	0	0	0

\* Denotes the subjects who used our tool.

Table 2.3: Debugging Study: Results of the subjects' time on each task (in minutes).

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8
P1	02:30	02:25	01:20	06:04	03:30	02:49	03:57	03:54
P2	04:41	07:35	02:45	09:52	04:18	03:00	03:33	03:07
P3	03:32	04:40	05:38	03:18	02:47	02:24	03:41	03:40
P4	02:52	02:13	03:45	05:54	03:13	01:38	03:34	03:01
P5 *	01:44	01:38	00:40	02:35	04:11	01:06	03:14	01:38
P6	03:00	02:10	02:43	02:43	04:02	02:00	02:39	02:18
P7 *	01:35	02:42	00:49	02:15	06:13	01:59	01:59	01:50
P8 *	01:52	03:56	00:47	02:10	04:44	01:12	01:49	01:58
P9	02:50	02:37	03:57	03:05	02:35	01:10	02:22	02:18
P10 *	01:58	03:06	01:08	03:33	04:47	01:56	03:34	01:55
P11 *	01:35	03:12	01:19	02:11	05:48	01:34	03:08	02:40
P12 *	00:58	02:00	00:42	03:29	03:25	01:00	02:22	01:24
P13 *	01:49	03:28	00:44	02:47	04:29	01:21	03:07	02:01
P14 *	01:12	02:13	00:37	02:44	04:31	01:21	03:20	01:47
P15 *	00:56	02:36	00:37	03:09	03:39	01:11	01:56	02:21
P16	02:24	04:59	01:55	09:01	03:12	02:44	03:00	03:42
P17	03:52	03:02	04:08	03:53	03:19	03:14	02:30	02:38
P18	04:52	02:32	04:11	03:20	02:28	02:26	02:34	04:10
geo-mean	03:18	03:15	03:07	04:43	03:13	02:17	03:02	03:08
median	03:00	02:40	01:19	03:14	03:51	01:47	03:04	02:20
stdev	00:55	01:25	01:36	02:16	01:02	00:43	00:39	00:50
geo-mean*	01:28	02:40	00:48	02:43	04:34	01:22	02:38	01:55
median*	01:35	02:42	00:44	02:44	04:31	01:21	03:07	01:55
stdev*	00:23	00:44	00:15	00:32	00:54	00:21	00:41	00:22

\* Denotes the subjects who used our tool.

Table 2.4: Debugging Study: Lostness results for the tasks for 18 subjects.

Participant	Task1		Task2		Task3		Task4		Task5		Task6		Task7		Task8	
	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L	R,N,S	L
P1	1,5,5	0.80	1,2,4	0.71	1,3,3	0.67	1,2,6	0.83	1,5,5	0.80	1,5,5	0.80	1,5,10	0.94	1,5,8	0.88
P2	1,5,5	0.80	1,4,10	0.96	1,5,10	0.94	1,5,14	1.03	1,5,14	1.03	1,5,6	0.82	1,5,6	0.82	1,5,10	0.94
P3	1,5,6	0.82	1,5,5	0.80	1,5,7	0.85	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80
P4	1,5,7	0.85	1,2,2	0.50	1,5,5	0.80	1,5,14	1.03	1,5,6	0.82	1,3,3	0.67	1,3,9	0.94	1,3,3	0.67
P5 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P6	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,5,8	0.88	1,5,5	0.80	1,5,7	0.85	1,5,5	0.80
P7 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P8 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P9	1,5,7	0.85	1,2,2	0.50	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,3,3	0.67	1,2,4	0.71	1,3,3	0.67
P10 *	1,1,1	0.00	1,1,1	0.00	1,2,2	0.50	1,2,2	0.50	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P11 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P12 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P13 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P14 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P15 *	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00	1,1,1	0.00
P16	1,1,1	0.00	1,5,5	0.80	1,5,5	0.80	1,4,4	0.75	1,5,5	0.80	1,5,5	0.80	1,5,6	0.82	1,5,8	0.88
P17	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,4,4	0.75	1,5,8	0.88	1,5,6	0.82	1,5,7	0.85	1,5,7	0.85
P18	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,5,7	0.85	1,5,5	0.80	1,5,5	0.80	1,5,5	0.80	1,5,8	0.88
mean		0.72		0.74		0.81		0.85		0.85		0.77		0.84		0.82
median		0.80		0.80		0.80		0.80		0.80		0.80		0.82		0.85
stdev		0.27		0.15		0.07		0.10		0.07		0.06		0.07		0.10
efficient subj. (< 0.5)		11.11%		0.00%		0.00%		0.00%		0.00%		0.00%		0.00%		0.00%
inefficient subj. (>= 0.5)		88.89%		100.00%		100.00%		100.00%		100.00%		100.00%		100.00%		100.00%
mean*		0.00		0.00		0.06		0.06		0.00		0.00		0.00		0.00
median*		0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.00
stdev*		0.00		0.00		0.17		0.17		0.00		0.00		0.00		0.00
efficient subj. (< 0.5)*		100.00%		100.00%		88.89%		88.89%		100.00%		100.00%		100.00%		100.00%
inefficient subj. (>= 0.5)*		0.00%		0.00%		11.11%		11.11%		0.00%		0.00%		0.00%		0.00%

\* Denotes the subjects who used our tool.



Table 2.5: Debugging Study: Post-session experience ratings.

Participant	Satisfaction (Q1)	Confidence	Capability (UMUX.Q1)	Usability (UMUX.Q2)
P1	5	5	5	5
P2	2	2	2	2
P3	2	2	2	1
P4	1	2	2	2
P5 *	6	6	7	7
P6	2	2	1	2
P7 *	7	7	7	6
P8 *	6	6	6	5
P9	1	5	2	3
P10 *	7	6	6	5
P11 *	6	6	6	7
P12 *	6	6	6	7
P13 *	6	7	6	6
P14 *	7	5	6	7
P15 *	6	7	6	6
P16	2	3	2	2
P17	2	2	1	2
P18	1	1	1	2
mean	2.00	2.67	2.00	2.33
median	2.00	2.00	2.00	2.00
stdev	1.22	1.41	1.22	1.12
mean*	6.33	6.22	6.22	6.22
median*	6.00	6.00	6.00	6.00
stdev*	0.50	0.67	0.44	0.83

\* Denotes the subjects who used our tool.

Table 2.6: Debugging Study: Pre-task expectation ratings.

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8
P1	1	3	1	3	3	3	4	2
P2	1	2	1	2	2	2	3	3
P3	1	1	1	1	1	1	1	3
P4	1	1	1	1	1	1	2	2
P5 *	2	2	2	4	2	2	2	3
P6	1	2	1	2	2	2	2	2
P7 *	3	4	3	4	5	5	4	5
P8 *	4	4	4	4	4	4	4	5
P9	2	2	2	3	2	2	3	4
P10 *	4	3	4	6	4	4	7	5
P11 *	2	5	2	5	4	4	4	4
P12 *	5	5	4	5	6	6	6	6
P13 *	2	4	3	5	4	4	4	3
P14 *	3	2	3	6	3	2	4	4
P15 *	2	2	2	3	4	4	3	5
P16	1	2	1	3	2	2	2	4
P17	1	2	1	2	2	2	2	2
P18	1	2	1	2	2	2	2	2
mean	1.11	1.89	1.11	2.11	1.89	1.89	2.33	2.67
median	1.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00
stdev	0.33	0.60	0.33	0.78	0.60	0.60	0.87	0.87
mean*	3.00	3.44	3.00	4.67	4.00	3.89	4.22	4.44
median*	3.00	4.00	3.00	5.00	4.00	4.00	4.00	5.00
stdev*	1.12	1.24	0.87	1.00	1.12	1.27	1.48	1.01

\* Denotes the subjects who used our tool.

Table 2.7: Debugging Study: Post-task experience ratings.

Participant	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8
P1	1	5	1	5	6	6	5	4
P2	4	4	1	4	6	7	7	7
P3	5	6	7	4	3	2	2	6
P4	2	1	3	4	2	2	3	3
P5 *	1	1	1	1	1	1	1	1
P6	3	4	5	4	5	4	4	4
P7 *	1	2	1	1	3	2	1	1
P8 *	1	2	1	3	3	1	2	2
P9	5	5	7	6	5	4	6	6
P10 *	2	1	2	4	2	1	2	1
P11 *	2	2	2	2	2	2	2	2
P12 *	1	2	1	2	3	1	2	1
P13 *	1	2	1	2	1	1	1	2
P14 *	1	1	1	1	2	1	1	1
P15 *	1	2	1	1	3	2	2	2
P16	2	5	5	4	5	5	6	7
P17	3	4	5	5	5	4	4	4
P18	4	4	4	4	5	5	4	4
mean	3.22	4.22	4.22	4.44	4.67	4.33	4.56	5.00
median	3.00	4.00	5.00	4.00	5.00	4.00	4.00	4.00
stdev	1.39	1.39	2.22	0.73	1.32	1.66	1.59	1.50
mean*	1.22	1.67	1.22	1.89	2.22	1.33	1.56	1.44
median*	1.00	2.00	1.00	2.00	2.00	1.00	2.00	1.00
stdev*	0.44	0.50	0.44	1.05	0.83	0.50	0.53	0.53

\* Denotes the subjects who used our tool.

Table 2.8: Debugging Study: Results of the disruption-related questions.

Participant	Expecting Disruption	Experienced Disruption (Q3)	Disruption-Worth Value (Q4)
P5	6	2	7
P7	4	1	7
P8	5	3	6
P10	7	2	7
P11	6	1	7
P12	6	1	6
P13	6	4	7
P14	6	1	7
P15	2	2	7
mean	5.33	1.89	6.78
median	6.00	2.00	7.00
stdev	1.50	1.05	0.44

Table 2.9: Summary results of verbal comments made during the Debugging study (part 1).

	Overall		Context		Satisfaction		Syntax		Recognition	
	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.
No. of Comments on UCAnDo-Models	29	30	2	2	3	16	1	1	0	2
No. of Comments on Other Tools	71	13	5	1	3	0	0	0	11	3
No. of Subjects who Commented on UCAnDoModels (Out of 9)	9	7	2	2	3	6	1	1	0	1
No. of Subjects who Commented on Other Tools (Out of 9)	9	4	2	1	2	0	0	0	4	3

Table 2.10: Summary results of verbal comments made during the Debugging study (part 2).

	Association		Action Planning		Confidence		Scalability		Intrusiv.	
	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.
No. of Comments on UCAnDo-Models	0	1	8	5	5	0	10	0	0	3
No. of Comments on Other Tools	19	4	17	4	15	1	1	0	0	0
No. of Subjects who Commented on UCAnDoModels (Out of 9)	0	1	5	4	3	0	5	0	0	2
No. of Subjects who Commented on Other Tools (Out of 9)	6	2	8	2	8	1	1	0	0	0

Table 2.11: Statistics related to analysed metrics in the Debugging user study.

Collected Metric	Tool	Mean	Median	Min	Max	Std. Dev.	<i>p-value</i>	<i>r</i>
Task Success	Our Tool	1.00	1.00	1.00	1.00	0.00	<b>0.0051</b>	<b>0.68</b>
	Other Tools	0.87	0.88	0.63	1.00	0.13		
Task Errors	Our Tool	0.00	0.00	0.00	0.00	0.00	<b>0.0005</b>	<b>0.82</b>
	Other Tools	0.47	0.38	0.00	1.25	0.38		
Task Time	Our Tool	2.32	2.31	1.92	2.74	0.28	<b>0.0010</b>	<b>0.78</b>
	Other Tools	3.42	3.27	2.61	5.04	0.74		
Lostness	Our Tool	0.01	0.00	0.00	0.13	0.04	<b>0.0002</b>	<b>0.88</b>
	Other Tools	0.80	0.81	0.71	0.92	0.06		
Expectation vs. Experience Ratings	Our Tool	-2.26	-2.25	-3.75	-1.38	0.74	<b>0.0003</b>	<b>0.84</b>
	Other Tools	2.46	2.50	1.25	3.13	0.64		
UMUX	Our Tool	6.22	6.50	5.50	7.00	0.51	<b>0.0003</b>	<b>0.88</b>
	Other Tools	2.17	2.00	1.50	5.00	1.12		

Table 2.11 presents the results of our quantitative analysis, in which we used the Mann-Whitney U test to assess the statistical significance of our empirical results from the Debugging study. As shown by the *p values* in the table (*p values*  $\leq 0.05$ ), the subjects who used UCAnDoModels had significantly better Task Success, fewer Errors on Tasks, shorter Times on Tasks, less Lostness, better Users' Experience versus Expectation ratings, and higher UMUX ratings (Usability Metric for User Experience) than the subjects who used other tools. Also, the effect size *r* indicates a *high practical significance* (*r*  $\geq 0.5$ ) between the subjects who used UCAnDoModels and the subjects who used other tools.