Online Appendix to: User-Experience from an Inference Perspective

PAUL VAN SCHAİK, Teesside University
MARC HASSENZAHΛ, Folkwang University
JONATHAN LING, University of Sunderland

APPENDIX B. PSYCHOMETRIC PROPERTIES OF MEASUREMENT INSTRUMENT (STUDY 1)

The results presented here are those from the analysis of Hassenzähl and Monk's [2007] basic inference model (see Figure 1), which includes all the psychometrically measured interaction-experience variables (pragmatic quality, hedonic quality, beauty and goodness). In testing the measurement model, reliability was analyzed (see Table II), and convergent and discriminant validity was assessed (see Table III). The reliability of each individual reflective item is assessed by its loading on the construct of which it is an indicator, which should be .70 or higher [Henseler et al. 2009]. More useful, in terms of interpretation of results, is Chin's [2010] recommendation to examine the squared loadings because each represents the proportion of overlap in variance between an item and any construct. Using a bootstrapping procedure, the loadings of all items were found to be statistically significant. Although Item PQ3's loading (after presentation only, but not after interaction) was below the cut-off point of .70, it was retained, given that it was significant and there was no cross-loading. At the construct level, reliability was analyzed using the composite reliability coefficient; all coefficients exceeded the cut-off point of .70.

Convergent validity (the extent of consistency among the items measuring a particular construct) was analyzed using the average variance extracted (AVE) by a construct from its indicators, which should be .70 or higher [Henseler et al., 2009]. The AVE values for pragmatic quality and hedonic quality after interaction exceeded or equaled this cut-off point; the values for pragmatic quality and hedonic quality after presentation only were .63 and .66, respectively, exceeding the value of .50. Therefore, on average, more variability in the items was accounted for by their respective factor than was not. Discriminant validity (the extent to which a measure of a particular construct differs from measures of other constructs) was assessed by analyzing the AVE by each construct from its indicators, which, according to the Fornell-Larcker-criterion, should be greater than its squared correlation with the remaining constructs. All values met this condition. The findings of discriminant validity provide evidence for the distinction between the dimensions of pragmatic quality and hedonic quality as separate indicators of experience. In this experiment and the following experiments, a composite score was created for each of the factors per participant, using the PLS weighted-average algorithm.

APPENDIX C. PSYCHOMETRIC PROPERTIES OF MEASUREMENT INSTRUMENT (STUDY 2)

In testing the measurement model, reliability was analyzed (see Table IV), and convergent and discriminant validity assessed (see Table V). Using a bootstrapping procedure, the loadings of all items were found to be statistically significant. Although Item PQ3's loading (after presentation only, but not after interaction) was below the cut-off point, it was retained, given that it was significant and there was no cross-loading. Some of the loadings of PQ items on the hedonic-quality factor were relatively

Table II. Coefficients of Reliability and Convergent Validity (Experiment 1)

(a) After presentation of Web site						
Construct/ indicator	Average variance extracted	Composite reliability	Loading		Standard error (main loading)	t^a (main loading)
			PQ	HQ		
Pragmatic quality	.63	.87				
- PQ1			.92	.40	.05	8.51
- PQ2			.88	.39	.04	1.23
- PQ3			.58	.18	.06	2.56
- PQ4			.76	.15	.05	4.61
Hedonic quality	.66	.88				
- HQI1			.36	.82	.04	7.66
- HQI2			.25	.74	.05	6.17
- HQI3			.39	.88	.03	1.04
- HQI4			.21	.80	.05	6.03

Note: PQ: pragmatic quality. HQ: hedonic quality. $p < .001$ for all values of t , except $p < .01$ for PQ3.

^aBootstrap, $N = 5000$

(b) After interaction with Web site						
Construct/ indicator	Average variance extracted	Composite reliability	Loading		Standard error (main loading)	t^a (main loading)
			PQ	HQ		
Pragmatic quality	.70	.90				
- PQ1			.90	.45	.02	36.29
- PQ2			.92	.52	.02	44.77
- PQ3			.76	.30	.08	9.79
- PQ4			.76	.33	.07	11.19
Hedonic quality	.80	.94				
- HQI1			.48	.89	.03	33.97
- HQI2			.45	.89	.03	35.75
- HQI3			.44	.90	.03	28.74
- HQI4			.40	.91	.02	41.85

^aBootstrap, $N = 5000$.

high (e.g., PQ1 and PQ2 in Table IV(a)). However, these results are still reasonable, given that these loadings were $< .70$ and therefore representing $< .50$ variance overlap, while the corresponding loadings on the pragmatic-quality factor were extremely high, $> .90$ and representing .83 to .86 variance overlap [Chin 2010]. At the construct level, reliability was analyzed using the composite reliability coefficient, which needs to be .70 or higher. All the coefficients exceeded this cut-off point.

Regarding convergent validity, the value for pragmatic quality after interaction exceeded the cut-off point of .70. The remaining values ranged from .63 to .69 (exceeding the value of .50), so more variance was accounted for than was not. Regarding discriminant validity, all values met the condition of AVE values greater than the squared correlations with the remaining constructs. In conclusion, the reliability, and the convergent and discriminant validity of the multi-item constructs were confirmed.

APPENDIX D. PSYCHOMETRIC PROPERTIES OF MEASUREMENT INSTRUMENT (STUDY 3)

In testing the measurement model, reliability was analyzed (see Table VI), and convergent and discriminant validity was assessed (see Table VII). Using a bootstrapping procedure, the loadings of the items for were found to be statistically significant. Although Item PQ3's loading (after presentation only, but not after interaction) was

Table III. Coefficients of Convergent and Discriminant Validity, and Descriptive Statistics (Experiment 1)

(a) After presentation of Web site				
	Beauty	Goodness	Hedonic quality	Pragmatic quality
Beauty	-			
Goodness	.27	-		
Hedonic quality	.28	.20	.66	
Pragmatic quality	.08	.39	.15	.63
Note: Off-diagonal values are squared correlations. Diagonal values are average extracted variances.				
(b) After interaction with Web site				
	Beauty	Goodness	Hedonic quality	Pragmatic quality
Beauty	-			
Goodness	.38	-		
Hedonic quality	.65	.48	.80	
Pragmatic quality	.09	.26	.24	.70
Note: Off-diagonal values are squared correlations. Diagonal values are average extracted variances.				
(c) Descriptive statistics				
	Beauty	Goodness	Hedonic quality	Pragmatic quality
Presentation-only				
- Mean	3.67	4.49	3.98	4.89
-SD	1.27	1.33	1.08	1.18
Hands-on experience				
- Mean	3.72	4.66	3.97	5.09
- SD	1.31	1.43	1.30	1.12
<i>d</i>	-0.01	0.19	0.12	0.05

d: standardized mean difference for the effect of interaction.

below the cut-off point of .70, it was retained, given that it was significant and there was no cross-loading. At the construct level, reliability was analyzed using the composite reliability coefficient; all coefficients exceeded the cut-off point of .70.

Regarding convergent validity, the AVE value for pragmatic quality after interaction exceeded the cut-off point of .70; the remaining values for pragmatic quality and hedonic quality ranged from .58 to .69, so more variance was accounted for than was not. Regarding discriminant validity, all values met the condition of AVE values greater than the squared correlations with the remaining constructs. In conclusion, the reliability, and the convergent and discriminant validity of the multi-item constructs were confirmed.

As a manipulation check, the effect of the independent variables task- and site complexity on task performance was tested. Performance was better when task and site were simple (see Table VIII). The significant effects task- and site complexity were very large and small to medium, respectively (see Table VIII).

Table IV. Coefficients of Reliability and Convergent Validity (Experiment 2)

(a) After presentation of Web site						
Construct/ indicator	Average variance extracted	Composite reliability	Loading		Standard error (main loading)	t^a (main loading)
			PQ	HQ		
Pragmatic quality	.63	.87				
- PQ1			.91	.67	.04	21.82
- PQ2			.93	.64	.03	27.97
- PQ3			.50	.13	.20	2.52
- PQ4			.77	.33	.11	7.26
Hedonic quality	.69	.90				
- HQI1			.48	.85	.85	16.95
- HQI2			.58	.91	.91	31.26
- HQI3			.51	.79	.79	9.04
- HQI4			.52	.77	.77	11.31

Note: PQ: pragmatic quality. HQ: hedonic quality. $p < .001$ for all values of t , except $p < .01$ for PQ3.

^aBootstrap, $N = 5000$.

(b) After interaction with Web site						
Construct/ indicator	Average variance extracted	Composite reliability	Loading		Standard error (main loading)	t^a (main loading)
			PQ	HQ		
Pragmatic quality	.78	.93				
- PQ1			.92	.50	.02	54.96
- PQ2			.94	.50	.02	6.25
- PQ3			.86	.42	.05	18.25
- PQ4			.80	.36	.06	14.53
Hedonic quality	.64	.88				
- HQI1			.47	.81	.06	14.62
- HQI2			.36	.80	.11	6.98
- HQI3			.39	.79	.10	7.76
- HQI4			.39	.80	.09	8.90

Note: PQ: pragmatic quality. HQ: hedonic quality. $p < .001$ for all values of t .

^aBootstrap, $N = 5000$.

Table V. Coefficients of Convergent and Discriminant Validity, and Descriptive Statistics (Experiment 1)

(a) After presentation of Web site				
	Beauty	Goodness	Hedonic quality	Pragmatic quality
Beauty	-			
Goodness	.36	-		
Hedonic quality	.58	.45	.69	
Pragmatic quality	.26	.66	.40	.63

Note: Off-diagonal values are squared correlations. Diagonal values are average extracted variances.

(b) After interaction with Web site				
	Beauty	Goodness	Hedonic quality	Pragmatic quality
Beauty	-			
Goodness	.35	-		
Hedonic quality	.38	.33	.64	
Pragmatic quality	.33	.70	.26	.78

Note: Off-diagonal values are squared correlations. Diagonal values are average extracted variances.

(c) Descriptive statistics				
	Beauty	Goodness	Hedonic quality	Pragmatic quality
After presentation				
- Mean	4.36	5.32	4.77	5.27
-SD	1.05	1.14	0.94	1.04
After interaction				
- Mean	4.11	3.97	4.28	3.80
- SD	0.98	1.45	0.97	1.49
<i>d</i>	-0.53	-0.90	-0.85	-0.25

d: standardized mean difference for the effect of interaction.

Table VI. Coefficients of Reliability and Convergent Validity (Experiment 3)

(a) After presentation of Web site						
Construct/ indicator	Average variance extracted	Composite reliability	Loading		Standard error (main loading)	t^a (main loading)
			PQ	HQ		
Pragmatic quality	.58	.84				
- PQ1			.91	.19	.05	***8.51
- PQ2			.88	.26	.04	***1.23
- PQ3			.58	.08	.06	**2.56
- PQ4			.76	.04	.05	***4.61
Hedonic quality	.64	.87				
- HQ11			.09	.82	.04	***7.66
- HQ12			.31	.74	.05	***6.17
- HQ13			.16	.88	.03	***1.04
- HQ14			.12	.80	.05	***6.03

Note: PQ: pragmatic quality. HQ: hedonic quality. * $p < .05$. ** $p < .01$. *** $p < .001$.

^aBootstrap, $N = 5000$

(b) After interaction with Web site						
Construct/ indicator	Average variance extracted	Composite reliability	Loading		Standard error (main loading)	t^a (main loading)
			PQ	HQ		
Pragmatic quality	.76	.93				
- PQ1			.92	.18	.02	55.14
- PQ2			.90	.14	.02	37.11
- PQ3			.81	-.02	.04	2.49
- PQ4			.85	.10	.04	23.31
Hedonic quality	.69	.90				
- HQ11			.14	.74	.07	1.92
- HQ12			.11	.85	.03	28.15
- HQ13			.07	.85	.04	19.30
- HQ14			.10	.89	.02	41.25

Note: PQ: pragmatic quality. HQ: hedonic quality. $p < .001$ for all values of t .

^aBootstrap, $N = 5000$.

Table VII. Coefficients of Convergent and Discriminant Validity, and Descriptive Statistics (Experiment 1)

(a) After presentation of Web site												
	Beauty			Goodness			Hedonic quality			Pragmatic quality		
Beauty	-											
Goodness	.31			-								
Hedonic quality	.52			.33			.64					
Pragmatic quality	.05			.28			.05			.58		
Note: Off-diagonal values are squared correlations. Diagonal values are average extracted variances.												
(b) After interaction with Web site												
	Beauty			Goodness			Hedonic quality			Pragmatic quality		
Beauty	-											
Goodness	.23			-								
Hedonic quality	.57			.23			.69					
Pragmatic quality	.01			.39			.02			.76		
Note: Off-diagonal values are squared correlations. Diagonal values are average extracted variances.												
(c) Descriptive statistics												
Site complexity	Beauty			Goodness			Hedonic quality			Pragmatic quality		
	Task complexity			Task complexity			Task complexity			Task complexity		
	Low	High	Overall	Low	High	Overall	Low	High	Overall	Low	High	Overall
After presentation												
Low												
- Mean	2.35	2.38	2.37	3.41	3.77	3.59	2.59	2.19	2.39	5.03	5.65	5.34
- SD	1.36	1.13	1.24	1.68	1.54	1.61	1.18	1.23	1.21	1.28	0.92	1.15
High												
- Mean	2.28	2.19	2.23	4.13	3.75	3.94	2.36	2.04	2.20	5.64	5.53	5.59
- SD	1.35	1.06	1.21	1.43	1.57	1.50	1.07	0.77	0.94	0.79	0.81	0.80
Overall												
- Mean	2.32	2.28	2.30	3.77	3.76	3.76	2.47	2.11	2.30	5.34	5.59	5.46
- SD	1.34	1.09	1.22	1.59	1.54	1.56	1.12	1.02	1.08	1.10	0.86	0.99
After interaction												
Low												
- Mean	2.59	2.23	2.41	4.00	3.10	3.56	2.63	2.14	2.39	4.87	4.06	4.47
- SD	1.21	1.12	1.17	1.92	1.66	1.84	1.19	1.03	1.13	1.38	1.67	1.57
High												
- Mean	2.47	1.94	2.20	4.44	2.91	3.67	2.43	1.96	2.19	5.21	3.70	4.46
- SD	1.22	0.98	1.13	1.32	1.53	1.61	1.07	0.83	0.98	0.96	1.29	1.36
Overall												
- Mean	2.53	2.08	2.31	4.22	3.00	3.61	2.53	2.05	2.29	5.04	3.88	4.46
- SD	1.21	1.05	1.15	1.65	1.59	1.72	1.13	0.93	1.06	1.19	1.49	1.46
<i>d</i>			0.01			-0.10			-0.01			-0.70

d: standardized mean difference for the effect of interaction.

Table VIII. Descriptive Statistics of Task Performance as a Function of Task- and Site Complexity, and Analysis of Variance (Experiment 3)

(a) Descriptive statistics						
Site complexity	Percentage of tasks completed			Percentage of tasks completed correctly		
	Task complexity			Task complexity		
	Low	High	Overall	Low	High	Overall
Low						
- Mean	44	29	37	40	24	32
- SD	12	11	14	13	11	14
High						
- Mean	40	21	30	35	12	24
- SD	13	10	15	13	7	16
Overall						
- Mean	42	25	33	38	18	28
- SD	13	11	15	13	11	16
(b) Analysis of variance						
Percentage of completed						
Source	<i>df</i>	Sum of squares	Mean square	<i>F</i>	<i>p</i>	ε^2
Site complexity	1	1219.55	1219.55	9.05	.003	.04
Task complexity	1	9387.91	9387.91	69.65	.000	.34
Site $\times$ task	1	169.84	169.84	1.26	.264	.00
Error	123	16578.58	134.785			
Total	126	27423.78				
Percentage of tasks completed correctly						
Source	<i>df</i>	Sum of squares	Mean square	<i>F</i>	<i>p</i>	ε^2
Site complexity	1	2284.93	2284.93	17.88	< .001	.07
Task complexity	1	12245.74	12245.74	95.83	< .001	.39
Site $\times$ task	1	408.08	408.08	3.19	.076	.01
Error	123	15717.84	127.787			
Total	126	30762.61				