

We had two groups of subjects: Group A first worked with colors, then switched to the colorless version. Group B started with the colorless version, then switched to the color version.

Correctness of subjects for static tasks. Since the χ^2 test requires expected frequencies to be larger than 5, we summarized static tasks.

Task	Group	Correct	Incorrect	χ^2	degrees of freedom	p value
S1+S2+S3	A	11	10	0	1	1.000
	B	11	10			
S4+S5+S6	A	13	8	0.104	1	0.747
	B	14	7			

Response time of subjects.

We make three comparisons of response time data. Hence, we have to adjust the significance level, e.g. with a Bonferoni correction. We divide the significance level (0.05) by 3, resulting in a significance level of 0.017 for a single comparison. This means that observed differences are significant, if the p value is not larger than 0.017.

We applied t-tests for independent and dependent samples. The response times are normally distributed, tested with a Kolmogorov-Smirnov-Test.

First: Between subjects:

Task	S1	S2	S3	M1	S4	S5	S6	M2
Mean A	3.04	5.28	4.07	13.32	4.54	6.52	4.10	24.45
Mean B	6.58	10.33	3.34	16.55	4.01	4.57	3.57	26.59
t value	-4.276	-2.759	0.919	-1.096	0.639	1.425	0.690	-0.282
p value	0.001	0.017	0.376	0.295	0.535	0.180	0.503	0.783

Second: Within subjects, group A

Task	S1/S4	S2/S5	S3/S6	M1/M2
t value	-2.117	-0.932	-0.042	-1.649
p value	0.079	0.387	0.968	0.150

Third: Within subjects, group B

Task	S1/S4	S2/S5	S3/S6	M1/M2
t value	3.973	3.591	-0.451	-2.770
p value	0.007	0.011	0.668	0.032

Opinion of subjects

We cannot compute exact p values, because our sample is too small. Instead, we look up whether the computed U value is significant in the table of U distribution for critical values

Difficulty

Task	S1	S2	S3	M1	S4	S5	S6	M2
U value	20.5	24.5	17.5	18	10.5	0	15.5	24
significant	no	no	no	no	yes	yes	no	no

Motivation

Task	S1	S2	S3	M1	S4	S5	S6	M2
U value	24	18.5	22	9.5	15.5	15	19.5	23
significant	no	no	no	yes	no	no	no	no

Performance with other version (worse or better)

Task	S1	S2	S3	M1	S4	S5	S6	M2
U value	6	2.5	0	13.5	2	3	2	4.5
significant	yes	yes	yes	no	yes	yes	yes	yes