UR
Universität Regensburg

Effects of different conditions for the translation of a test on the PC

Ludwig Kreuzpointner
Institute for Psychology, Universität Regensburg

Introduction

The ITC Guidelines on Computer-Based and Internet-delivered Testing (2005) listed the requirements when a CBT is developed from a PBT.
1. clear documented evidence of the equivalence: same means and sd, high correlation, comparable reliabilities and validities.
2. equivalent test taking in item control (e.g. to skip or review items), item presentation and the format for responding
3. equivalent test conditions especially for internet testing

The control of equivalence is the observance of the first point. When PBT and CBT are psychometric parallel tests they can be seen as equivalent. So we decide to collate the PBT with several different kind of presentation on the pc to identify a best implementation to anticipate equivalence.

Design

In search of an optimal design of the computerized presentation we vary three factors: *presentation*, *answer* and *time*. Either the items were presented one by one or they were shown altogether like on the paper sheet (*presentation*). In *answer* we differ the opportunities to change the given answer either the whole time of working the subtest or just until the next item is shown or the condition that no change is possible after denoting. Each subtest has a time limit. In the factor *time* we vary the conditions that the subject has or has not information about the actual time, and also a third condition with a time limit for each item. Each participant edits each of the 11 subtests in another random assigned condition (see table 1). The used test battery (revised Leistungsprüfsystem, LPS-neu; Kreuzpointner, 2010; originally Horn, 1983) is constructed as follows: The task of subtest 1 is to find the wrong letter of a misspelled word. In subtest 2 an anagram is to solved. The subtests 3 to 5 are figural, numerical and alphanumerical sequences with a mistake to identify. At subtest 6 a mirror-inverted sign is to find. The task of subtest 7 is to count the sides of 3D-objects. At subtest 8 one of five patterns has to find within complex pattern. On subtest 9 a column of numbers and letters has to scan and marking each eight 0, than each eight 1 and so on. For subtest 10 two identical rows has to scan for changes. The task of subtest 11 is to add up ten digits.

Table 1: Conditions and experimental design

Presentation	one by one									altogether					
Answer	once			until next			all the time			once		until next		all the time	
Time	no	global	single	no	global	single	no	global	single	no	global	no	global	no	global
Condition	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Study

N=205 pupils of a vocational school completed the full test in both modi within 8 weeks; 107 first the CBT; 98 first the PBT of the new LPS. In order to find the best condition for equivalence, the analyses do not examine the interaction effects of the three modus-factors with each other. Ideally there should be a main effect neither for the modus (PBT or CBT, see figure 2) nor for the order of presentation (PBT at T1 or PBT at T2). Since a retest-effect cannot excluded, an interaction is expected the kind of a diagonal cross, with the first performance of each modi on the same level just as the second performance just a bit higher.

Results

Figure 1 illustrates that there are conditions of the CBT that leads to same means as the PBT, especially for the reasoning tasks of subtest 3, 4 and 5, but also 2, 9 and 11. The easier the items are, the higher is the mean of the PBT (see subtests 6, 7, 8, and particularly 10). Subtest 1 exemplifies the effect that the degree of unfamiliarity with the tasks and likewise with the test itself comes to lower performance especially at the CBT. In figure 2 this effect can be seen in the higher difference of CBT at T1 to PBT at T2 (within comparison) as from PBT at T1 to CBT at T2 and the higher performance at PBT at T1 compared with the CBT at T1 (between comparison). In table 2 the results of ANOVAs for each subtests is shown as η^2. Except the main-effect of the modus of the subtests 3 and 4 all effects are statistical significant ($\alpha = 5\%$). Overall there is a multivariate interaction-effect (Pillai $V = 0.76$, $F(11, 193) = 56.71$, $p < .01$, $\eta_p^2 = .76$). The main-effects of the order ($V = 0.15$, $F(11, 193) = 3.16$, $p < .01$, $\eta_p^2 = .15$) and the modus ($V = 0.63$, $F(11, 193) = 30.20$, $p < .01$, $\eta_p^2 = .63$) are multivariate statistical significant, too.

The single comparison of PBT and CBT of each subtest in each condition is presented in table 3 in form of Cohen's d'. Yellow fields mark small effects, green fields label almost none effects. Hedges & Olkin (1985) introduced a meta-analytic method to summarize these effects meaningful when homogeneity is given (here $Q < 13.99$). By excluding subtest 9 and 10 most condition fulfills this criterion. The resulting mean effect portends a better mean performance in the PBT.

In table 4 the Pearson correlations between the PBT and the CBT and the Cronbach's alpha of the PBT is shown. There are some good and some acceptable values (with provision for the small N of each field; $12 < N < 31$). The mean correlation is not as quit as good as it should be ($.45 < r < .62$).

Discussion

The interactions of modus and working sequence indicate for the most subtests that an unknown test is more difficult as CBT than as PBT. So a first general act to boost equivalence is to increase familiarity with the test principles. There are some tests which are more robust against different variation of condition than other. One main different is the difficulty of the items. The easier the items the more problematic is the translation on the PC (see subtest 10). The investigated conditions of item presentation, answer modality and time presentation do not lead to a clear answer, whether combination should use. Condition 4 and 5, which are used quite often when translating tests, seem to be an eligible choice when single item presentation is required. Conditions 7 and 8, when subjects can go back to worked items via a list of all items on the right side by working on a highlighted item in the center could be the best compromise when need single item presentation, but the need of equivalence, too. The putative same presentation of PBT as CBT, condition14 and 15 leads just partly to equivalence result, so the opportunities of single item presentation should be used without qualm. Sadly we do not find a golden rule for translating a given PBT. But even if we had found one, the documentation of equivalence must run for each test ever. It's hard but it's work!

References

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
Horn, W. (1962, 1983). *Leistungsprüfsystem L-P-S*. Göttingen: Hogrefe.
ITC, International Test Commission (2005). *International Guidelines on computer-based and Internet delivered testing*. (http://www.intestcom.org/guidelines).
Kreuzpointner, L. (2010). *Bedingungen für die Äquivalenz von Papier-Bleistift-Version und Computerversion bei Leistungstests*. Regensburg: Universitätsverlag.

Figure 1: Mean performance in the serveral conditions in each subtest, the PBT and the overall CBT
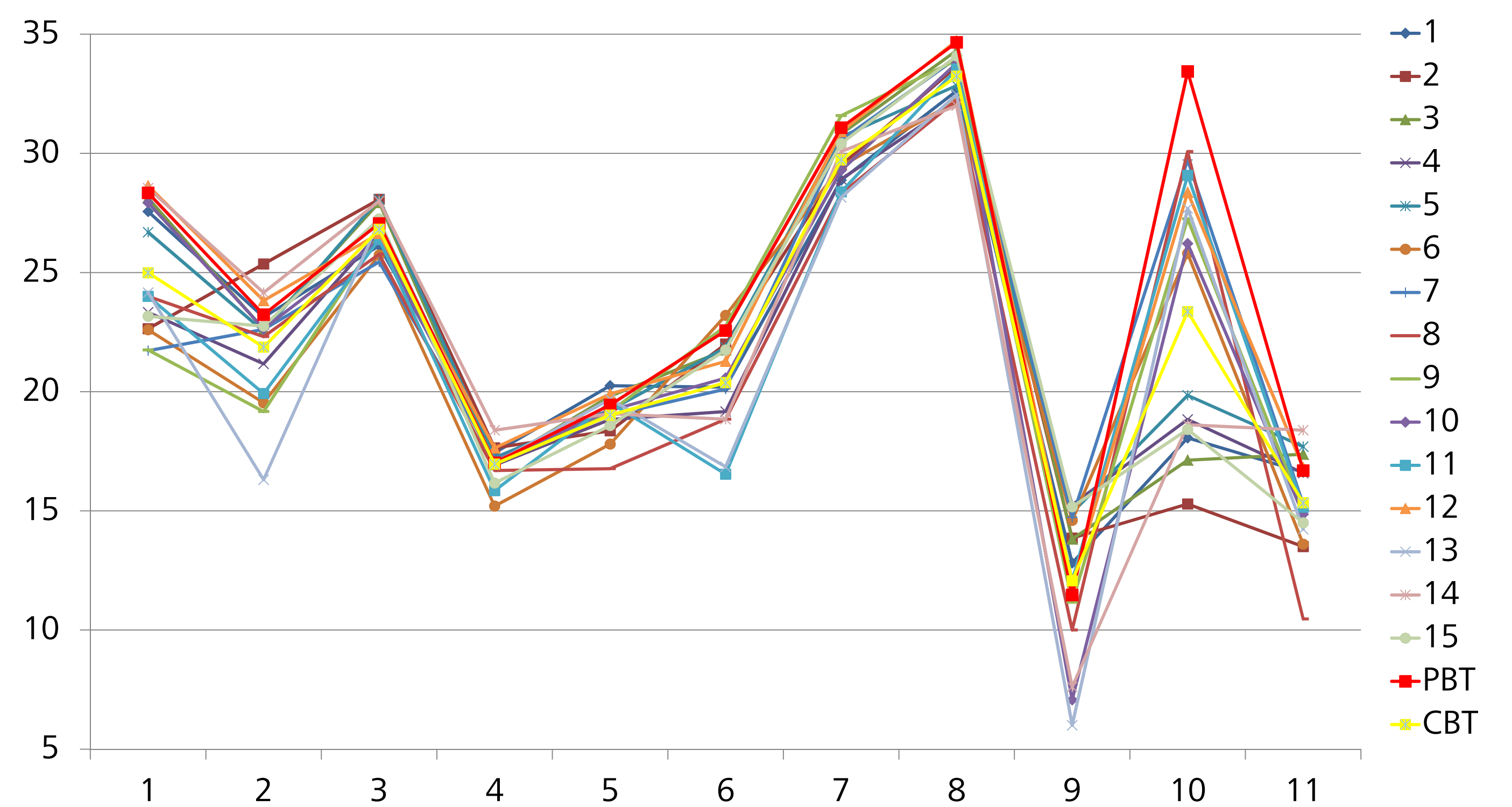

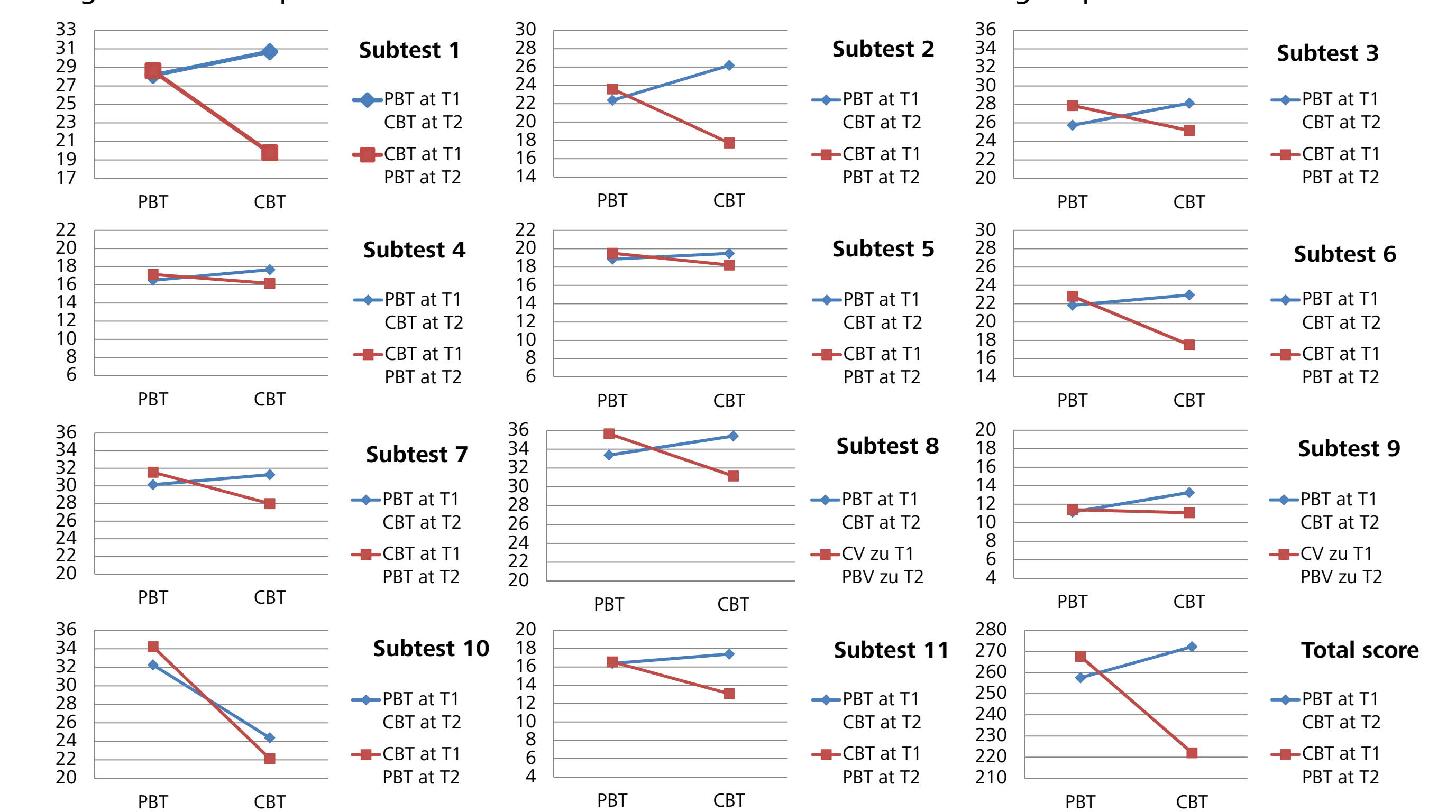Figure 2: Mean performance of the PBT and the CBT in both working sequences

Table 2: η^2 of the within effects "modus" and the interaction of modus and working sequence

Effect	Subtest											M	SD
	1	2	3	4	5	6	7	8	9	10	11		
Modus	.162	.043	.004	.002	.023	.139	.073	.091	.067	.676	.048	.121	.191
Modus × PBT-CBT/CBT-PBT	.378	.494	.369	.151	.102	.259	.238	.333	.064	.074	.137	.236	.143

Table 3: Cohen's d' of the comparison PBT with CBT regardless of the working sequence, statistical means, sd, teststatistic of the homogeneity test by Hedges & Olkin (1985) and the consequential d' by provision of the intercorrelation

	Condition	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	-0.32	-0.28	-0.30	-0.46	-0.18	-0.57	-1.33	-0.39	-0.74	-0.46	-0.37	-0.53	-0.36	-0.72	-0.63
	2	-0.28	-0.40	-0.06	-0.59	-0.11	-0.28	-0.09	0.06	-0.20	-0.32	-0.25	-0.51	-0.68	-0.50	0.00
	3	-0.20	0.71	-0.11	-0.14	-0.35	-0.15	-0.03	0.11	-0.56	-1.11	0.00	-0.14	-0.34	-0.34	-0.47
	4	0.05	0.04	-0.25	0.00	-0.18	-0.10	-0.05	-0.02	-0.83	0.50	0.13	0.24	0.00	0.00	0.04
	5	-0.28	-0.07	-0.40	-0.51	0.27	-0.58	-0.40	0.20	-0.25	-0.18	0.22	-0.77	-0.21	-0.15	0.16
	6	-0.01	-0.44	-0.66	-0.24	-1.13	-0.68	-0.68	-0.55	-0.82	-0.56	-0.44	-0.66	-0.33	-0.21	-0.23
	7	-0.19	-0.42	-0.38	-0.69	-0.38	0.11	-0.99	-0.68	-0.78	-0.21	-0.06	-0.41	-0.62	-0.74	-0.17
	8	0.13	-0.08	-0.33	-0.90	-0.36	-0.91	-0.22	-0.35	-0.71	-0.24	-0.29	-0.27	-0.25	-0.55	-0.62
	9	1.54	0.05	-	-0.16	-0.25	-	-1.49	-1.01	-	-	-	1.58	0.76	1.27	1.26
	10	-0.52	-4.95	-	-1.60	-0.63	-	-1.48	-0.57	-	-	-	-4.45	-4.12	-4.95	-5.03
	11	-0.39	-0.36	0.04	0.06	-0.30	-0.08	-0.93	-0.10	-0.43	-0.34	-0.30	-0.28	0.10	-0.63	-0.09
Parameter	M(d)	-0.04	-0.56		-0.48	-0.33		-0.70	-0.30				-0.56	-0.53	-0.68	-0.53
	SD(d)	0.56	1.49		0.48	0.35		0.58	0.38				1.44	1.25	1.52	1.58
	Q	28.49	72.10		25.74	13.91		41.70	16.52				95.03	128.1	83.36	137.4
	d'	0.02	0.26		-0.37	-0.43		-1.07	-0.48				0.60	0.12	0.47	0.16
Parameter \ (9;10)	M(d)	-0.17	-0.14	-0.27	-0.39	-0.30	-0.35	-0.66	-0.52	-0.19	-0.59	-0.32	-0.15	-0.37	-0.28	-0.43	-0.22
	SD(d)	0.18	0.36	0.21	0.32	0.37	0.34	0.48	0.31	0.25	0.42	0.23	0.30	0.26	0.26	0.29
	Q	2.82	10.96	4.45	9.43	13.11	13.16	21.31	8.88	7.14	21.95	5.06	10.28	6.39	6.35	8.35
	d'	-0.15	-0.15	-0.27	-0.32	-0.37	-0.33	-0.59	-0.25	-0.68	-0.25	-0.21	-0.31	-0.24	-0.45	-0.34

Table 4: Pearson's r of the comparison PBT with CBT regardless of the working sequence, statistical means, sd, teststatistic of the homogeneity test by Hedges & Olkin (1985) and the consequential d' by provision of the intercorrelation

	Condition	PBT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	.92	.73	.46	.74	.29	.66	.40	.62	.78	.46	.47	.74	.54	.82	.58	.08
	2	.90	.43	.70	.61	.33	.59	.48	.76	.69	.78	.60	.73	.54	-.59	.65	.64
	3	.76	.53	-.26	.31	.45	.45	.62	.25	.45	.51	.53	.34	.72	-.11	.30	.08
	4	.65	.80	.56	.11	.17	.38	.85	.34	.73	.12	.65	.12	.64	.75	.66	.02
	5	.74	.46	.45	.44	.64	.57	.44	.64	.78	.54	.01	.80	.64	.43	.44	.65
	6	.93	.43	.78	.73	.86	.64	.51	.68	.78	.48	.47	.31	.31	.79	.32	.71
	7	.86	.32	.76	.75	.44	.07	.69	.28	.58	.57	.71	.77	.54	.86	.35	.77
	8	.90	.45	.52	.73	-.02	.70	.35	.71	.30	.77	.69	.29	.47	.45	.70	.16
	9	.80	.20	.41	-	.44	.32	-	.26	.17	-	-	-	.41	.52	.59	.25
	10	.93	.38	.29	-	.37	.28	-	.64	.28	-	-	-	.57	.75	.71	.45
	11	.91	.55	.83	.83	.67	.38	.48	.38	.80	.57	.63	.74	.64	.62	.92	.64
Parameter	Q		9.49	20.91	13.50	16.86	9.17	8.68	12.55	18.01	7.93	7.24	15.4	4.47	41.9	13.4	17.8
	z		.55	.61	.72	.50	.49	.64	.58	.67	.62	.61	.67	.61	.70	.68	.49
	r'		.50	.54	.62	.46	.45	.56	.52	.59	.55	.55	.59	.54	.60	.59	.45
	u_r		.39	.44	.51	.35	.35	.45	.43	.49	.43	.43	.48	.44	.52	.50	.34
	o_r		.60	.63	.70	.56	.55	.65	.61	.67	.65	.65	.68	.63	.68	.67	.55