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 documentation 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 the 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 offer the differences 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 scan and marking each eight 0, then each eight 1 and so on. For subtest 10 two identical rows has to scan for solved. The subtests 3 to 5 are figural, numerical and alphanumerical sequences with a mistake to identify. At the task of subtest 1 is to find the wrong letter of a misspelled word. In subtest 2 an anagram is to be solved. There exists 3 to 5 are figural, numerical 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 be scanned until a number or letter is right 0, than change 1 and on. For subtest 10 identical rows has to scan for changes. The task of subtest 11 is to add up digits.

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 UPS. In order to find the best condition for equivalence, the analyses do not examine the test performance of the three test modus but compare the 15 conditions 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 modus on the same level just as the second performance just a bit higher.

Results

In figure 1 it 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 becomes 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 well as between conditions (PBT at T2 vs. CBT at T2). Since a retest-effect cannot excluded, an interaction is expected the kind of a diagonal cross, with the first performance of each test on the same level just as the second performance just a bit higher.

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 materials. There are some tests which are more robust against different variation of condition than other. One main difference of the given items: the visual input sequences with the condition of 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 in other conditions, are multivariate statistical significant, too.

References


Poster presented at the 7th Conference of the International Test Commission at the Chinese University of Hong Kong in Hong Kong. China. ludwig.kreuzpointner@uni-regensburg.de

The presentation is sponsored by the "Frieden der Kommunikation e.V."