

Does Computerised Test Administration Change the Nature of the Test?

Stefan Bondorowicz PhD

Research psychologist, Psytech International.

The rapid development in computer technology as led to changes in the nature of tests & testing. Computer software packages are now available which can administer even the most complicated of tests and then score these tests as responses are made to the test items. One important aspect of the change from paper & pencil to computer-based administration, concerns the effects the change in test format might have on the available normative data for a test. In other words does the translation of a test to computer format alter the nature of the test itself? It is possible that the score range of a test administered in question booklet form may be different from the range when the test is administered via a computer. Given that most of the normative data available for mainstream psychometric tests was collected from paper & pencil test administration the question is far from academic. For instance, many organisations will administer computer-based tests at their head office location but will use paper & pencil format when their testers visit remote locations. If comparison is to be made across a group in which some individuals received paper & pencil testing and some computer testing then the importance of the question of normative comparability can not be overstated.

Very little research attention has been paid to this topic, which is surprising given the possible impact that format differences would have. Roid (1986) has indicated that what little evidence is available concerning paper & pencil v. computer formats suggests that computer administration of most tests does not change the score range enough to affect the normative basis of the test. While this is somewhat reassuring the number of studies looked at by Roid was small and consisted entirely of American investigations.

This paper examines the effect of administration format on both reasoning and personality test performance. The Reasoning test chosen was the Graduate Reasoning Test, a graduate level test comprising three subtests - verbal, numerical and abstract. This test was chosen as being representative of many mainstream reasoning tests. The Graduate Reasoning Test also gives an opportunity to test whether the representational format of a test is important. It could be the case that computer presentation of graphical material, such as is found in mechanical & spatial reasoning tests, might lead to performance differences while alphanumeric material does not. With the GRT the abstract subtest uses a graphical format while the verbal and numerical

subscales use alphanumeric formats. The personality test used was the 15 Factor Questionnaire (1991), a 191 item trait based measure, which is fairly typical of personality questionnaires in widespread use in UK industry and commerce.

The Study

A group of 120 university undergraduates took part in this investigation. The students were divided into four groups. Each group completed the Graduate Reasoning Test twice with an interval of two weeks between successive administrations. Two of the groups completed the same format of the test on each occasion - ie paper/paper or computer/computer. The other two groups experienced both formats of the test, each group in a different order - i.e. paper/computer or computer/paper.

Results

An independent t-test was used to test whether any differences existed between the mean scores on first administration for those students who completed the paper version of the GRT1 and those receiving the computer administered version. As can be seen from Table 1 a significant difference was found for both the Numerical and Abstract sub-scales in that students who had received the computer administered version of the test tended to score significantly higher than those who had received the paper version first. Table 2 provides similar data for the two test formats on second administration. It can be seen from this table that no significant differences existed between the different formats for second administration.

A similar analysis was completed with the personality measure simply comparing mean scale score differences and as can be seen from Table

Conclusion

The results of this study show that as far as the Graduate Reasoning Test, at least, is concerned the format in which the test is administered does affect, to some extent, the scores obtained. No differences between the group means were detected for either of the three subtests of the Graduate Reasoning Test. It was also the case that no difference was found between the graphical representation of the abstract subtest and the

alphanumeric representation of the verbal & numerical subtests. Furthermore no interaction effects were observed which indicates that the well established phenomenon of 'practice effects' does not differ with the nature of the test medium.

These results provide some confidence that the changeover from paper and pencil to computer-based testing will not require the restandardisation of tests. It would seem from this study that administration of ability tests by either computer or paper & pencil will produce similar performance levels. Thus it is perfectly acceptable to compare individuals who were tested using different formats.

Dr Stefan Bondorowicz is a research psychologist with Psytech International.